

ANGLIA RUSKIN UNIVERSITY

AN EMPIRICAL STUDY TO ASSESS THE IMPACT OF MOBILE TOUCH-SCREEN
LEARNING ON USER INFORMATION LOAD

JOHN TALBOT

A thesis in partial fulfilment of the requirements of Anglia Ruskin University
for the degree of Doctor of Education.

Submitted: June 2017

ACKNOWLEDGEMENTS

On the day I started this doctorate I opened a fortune cookie at the course dinner. My fortune read *"Your path is arduous but will be amply rewarding"*. Thanks to the following people, my path was not arduous but was, for the most part, a pleasant and illuminating journey.

I would like to thank Geraldine Davis for accepting me on the course and for her kind advice throughout.

I would like to thank my supervisors Debbie Holley, Philip Howlett and Andy McVicar for their critical evaluation of the work in progress and ongoing support.

I would like to thank Lauren Godier for her very thorough critical-read and Jufen Zhang for confirming that the statistical tests had been used appropriately and correctly.

ANGLIA RUSKIN UNIVERSITY

ABSTRACT

FACULTY OF EDUCATION

DOCTOR OF EDUCATION

AN EMPIRICAL STUDY TO ASSESS THE IMPACT OF MOBILE
TOUCH-SCREEN LEARNING ON USER INFORMATION LOAD

JOHN TALBOT

June 2017

This study evaluated cognitive task load imposed on adult mobile-learners studying the subject of human anatomy. Touch-screen computers were originally conceptualised as a replacement for textbook learning and there are electronic versions of many popular anatomy books available for smartphones and tablets. However, Human Computer Interaction is known to increase cognitive load in the user, which could present a barrier to learning. To date there have been no empirical studies performed to measure differences in cognitive load between mobile-learners and textbook learners, particularly using a representative demographic for distance-learners.

The research was designed to answer the following question: Is there a statistically significant difference in the level of task load experienced by a learner when undertaking an interactive multimedia learning activity delivered by a mobile touch-screen device compared to that experienced by a learner undertaking an equivalent non-interactive learning activity?

Cognitive load is quantifiable, so the study employed a cross-sectional, experimental, two-armed controlled trial to measure and compare differences in levels of self-reported task load between two parallel, balanced groups of learners during a learning activity. The learning task was to memorise the foramina and associated structures found in the human skull-base. The NASA Task Load Index was used to measure six dimensions of task load namely; mental, physical and temporal demand, performance, effort and frustration. The experimental group used mobile devices, the control group studied a labelled photograph. The results of the study provide an original contribution to knowledge in that they demonstrate that smartphone-learners experienced a significantly lower task load than non-interactive learners, whereas tablet-learners did not. Mobile-learners reported significantly lower levels of mental demand and effort than non-interactive learners. Smartphone learners reported significantly lower levels of net task load, mental demand, physical demand and effort than tablet-learners. It was concluded that effective use of multimedia may ameliorate any cognitive load placed on users by the device and that smartphones may provide a better platform for this type of m-learning than tablet computers.

Key Words:

Mobile-learning, smartphones, tablets, cognitive load theory, human computer interaction

Contents

1. CHAPTER 01: INTRODUCTION	1
1.1. RESEARCH AIM.....	2
1.2. RESEARCH OBJECTIVES.....	3
1.3. A NOTE ON WRITING STYLE.....	4
1.4. PROFESSIONAL BACKGROUND.....	5
1.4.1. <i>Professional Practice: Clinical Experience</i>	6
1.4.2. <i>Professional Practice: Computer-Based Learning</i>	8
1.4.3. <i>Professional Practice: Scholarly Activity</i>	12
1.5. CONCEPTUAL AND THEORETICAL FRAMEWORKS.....	14
1.5.1. <i>Educational Theory and m-learning</i>	14
1.5.2. <i>Mobile Learning Through a Cognitivist lens</i>	16
1.5.3. <i>Cognitive Load Theory</i>	22
1.5.4. <i>Human-Computer Interaction</i>	23
1.6. METHODOLOGICAL CONSIDERATIONS	26
1.6.1. <i>Ontology</i>	27
1.6.2. <i>Epistemology</i>	27
1.7. MOBILE LEARNING AS A RESEARCH TOPIC	30
1.7.1. <i>Touch-screen Devices in the Context of e-learning</i>	34
1.7.2. <i>The Benefits of m-learning</i>	37
1.7.3. <i>The Limitations of m-learning</i>	41
1.8. A STATEMENT OF THE RESEARCH PROBLEM.....	45
1.9. RESEARCH QUESTIONS – THE PURPOSE OF THIS STUDY	47
1.10. STATEMENT OF RESEARCH HYPOTHESES.....	48
1.10.1. <i>Primary Hypotheses</i>	48
1.10.2. <i>Assumptions</i>	48
1.11. SUMMARY OF CHAPTER 01	50
2. CHAPTER 02: LITERATURE REVIEW.....	52
2.1. INTRODUCTION.....	52
2.2. COGNITIVE LOAD THEORY IN INSTRUCTIONAL DESIGN	60
2.2.1. <i>Intrinsic Cognitive Load</i>	62
2.2.2. <i>Extraneous Cognitive Load</i>	62
2.2.3. <i>Germane Resources (Germane Cognitive Load)</i>	64
2.3. CLT AND LEARNING	65
2.3.1. <i>Goal-Free Effect</i>	65
2.3.2. <i>Worked Example Effect</i>	66
2.3.3. <i>Split-Attention Effect</i>	66
2.3.4. <i>Modality Effect</i>	66
2.3.5. <i>Redundancy Effect</i>	67

2.4. CRITICISMS OF COGNITIVE LOAD THEORY	69
2.4.1. <i>Conceptual Criticisms of CLT</i>	70
2.4.2. <i>Methodological Criticisms of CLT</i>	71
2.4.3. <i>Ontological Criticism of CLT</i>	72
2.5. THE COGNITIVE THEORY OF MULTIMEDIA.....	75
2.6. CLT AND M-LEARNING	76
2.6.1. <i>Device Usability</i>	77
2.6.2. <i>Situated Learning</i>	81
2.6.3. <i>Material design and mode of presentation</i>	88
2.6.4. <i>Engagement and Adoption</i>	92
2.7. LITERATURE-BASED JUSTIFICATION FOR THIS STUDY	96
2.7.1. <i>Explicit Recommendations for Further Research Made in The Literature</i>	97
2.8. SUMMARY OF CHAPTER 02	101
3. CHAPTER 03: METHODOLOGY.....	103
3.1. INTRODUCTION.....	103
3.2. DESIGN OVERVIEW.....	103
3.3. METHODS CONSIDERED FOR MEASUREMENT OF COGNITIVE LOAD	105
3.3.1. <i>Direct, Objective Measurement</i>	106
3.3.2. <i>Indirect, Objective Measurement</i>	110
3.3.3. <i>Direct, Subjective Measurement</i>	112
3.3.4. <i>Indirect, Subjective Measurement</i>	113
3.4. DESIGN RATIONALE	115
3.4.1. <i>Independent Variables</i>	116
3.4.2. <i>Dependent Variables</i>	116
3.5. INSTRUMENTATION USED IN THIS STUDY.....	117
3.5.1. <i>The NASA Task Load Index</i>	117
3.5.2. <i>Rationale for Using NASA TLX in this Study</i>	119
3.5.3. <i>Pre and Post-Testing</i>	120
3.5.4. <i>Rationale for Using Pre/Post-Testing in This Study</i>	121
3.6. RESEARCH HYPOTHESES.....	122
3.6.1. <i>Between Groups</i>	122
3.6.2. <i>Between Smartphone Learners and Control Group</i>	122
3.6.3. <i>Between Tablet PC Learners and Control Group</i>	123
3.6.4. <i>Between Participants Grouped by Device type used (Smartphone or Tablet)</i>	123
3.6.5. <i>Between Participants Grouped by Screen Size of Device Used</i>	124
3.6.6. <i>Between Participants Grouped by Spatial Resolution of Device Display</i>	124
3.7. SAMPLING	126
3.7.1. <i>Determining Participant Numbers for Statistical Power</i>	126
3.7.2. <i>Sampling Method</i>	127
3.7.3. <i>Participant Demographics</i>	128
3.7.4. <i>Inclusion Criteria</i>	129
3.7.5. <i>Exclusion Criteria</i>	129

3.7.6.	<i>Ethical Approval</i>	129
3.7.7.	<i>Informed Consent</i>	130
3.7.8.	<i>Confidentiality</i>	131
3.8.	LEARNING MATERIAL AND DATA COLLECTION DESIGN	132
3.8.1.	<i>Originally Proposed Data Collection Method (Discontinued)</i>	133
3.8.2.	<i>Revised Data Collection Method (Used in this Study)</i>	135
3.9.	DATA COLLECTION PROCEDURE.....	142
3.9.1.	<i>Stage 01: Participant Information and Consent</i>	142
3.9.2.	<i>Stage 02: App installation</i>	142
3.9.3.	<i>Stage 03: Pre-Test</i>	142
3.9.4.	<i>Stage 04: Learning Activity</i>	143
3.9.5.	<i>Stage 05: NASA TLX Questionnaire</i>	145
3.9.6.	<i>Stage 06: Post Test</i>	145
3.10.	METHOD OF DATA ANALYSIS	147
3.10.1.	<i>Descriptive Statistics</i>	147
3.10.2.	<i>Inferential Statistical Tests</i>	147
3.10.3.	<i>Rationale For the Statistical Tests Used</i>	147
3.10.4.	<i>Effect Size Calculations</i>	153
3.10.5.	<i>Summary of Decision-Making Process for Statistical Tests Used</i>	155
3.10.6.	<i>Methodological Limitations</i>	156
3.10.7.	<i>Delimitations</i>	157
3.11.	SUMMARY OF CHAPTER 03	158
4.	CHAPTER 04 – RESULTS	159
4.1.	DISTRIBUTION OF DATA	159
4.1.1.	<i>Distribution of NASA TLX Weighted Scores Q-Q Plots</i>	161
4.1.2.	<i>Distribution of Pre/Post Test Scores (Marks),</i>	165
4.1.3.	<i>Distribution of Data from the Subscales of the NASA TLX</i>	167
4.2.	SOFTWARE USED FOR STATISTICAL ANALYSIS	169
4.3.	MARKING PRE AND POST-TEST SCORES	170
4.3.1.	<i>Marking Criteria for Pre and Post-Testing</i>	171
4.4.	PARTICIPANT DEMOGRAPHICS	172
4.4.1.	<i>Experimental Group</i>	172
4.4.2.	<i>Control Group</i>	173
4.5.	PRINCIPAL FINDINGS.....	174
4.5.1.	<i>Comparing Net Task load Between Groups (H1)</i>	174
4.5.2.	<i>Comparing Test Scores Between Groups (H2)</i>	176
4.5.3.	<i>Comparing Mental Demand Between Groups (H3)</i>	178
4.5.4.	<i>Comparing Physical Demand Between Groups (H4)</i>	179
4.5.5.	<i>Comparing Temporal Demand Between Groups (H5)</i>	180
4.5.6.	<i>Comparing Performance Between Groups (H6)</i>	182
4.5.7.	<i>Comparing Effort Between Groups (H7)</i>	184
4.5.8.	<i>Comparing Frustration Between Groups (H8)</i>	186
4.5.9.	<i>Summary of Results for Comparisons Between Groups</i>	187

4.6. SIGNIFICANT RESULTS BETWEEN SMARTPHONE-LEARNERS (N=36) AND CONTROL GROUP (N=65).....	188
4.6.1. Comparing Net Task load Between Smartphone and Control Groups (H9)	188
4.6.2. Comparing Mental Demand Between Smartphone and Control Groups (H11)	190
4.6.3. Comparing Effort Between Smartphone and Control Groups (H15)	192
4.7. SIGNIFICANT RESULTS BETWEEN TABLET-LEARNERS (N=24) AND CONTROL GROUP (N=65).....	193
4.7.1. Comparing Physical Demand Between Tablet and Control Groups (H20)	193
4.8. SECONDARY FINDINGS: RESULTS RELATING TO DEVICE TYPE.....	194
4.8.1. Comparing Net Task load Between Learners Grouped by Device Type (H25).....	195
4.8.2. Comparing Test Scores between Learners Grouped by Device Type (H26).....	197
4.8.3. Comparing Mental Demand Between Learners Grouped by Device Type (H27).....	198
4.8.4. Comparing Physical Demand Between Learners Grouped by Device Type (H28)	199
4.8.5. Comparing Temporal Demand Between Learners Grouped by Device Type (H29)...	200
4.8.6. Comparing Performance Between Learners Grouped by Device Type (H30).....	201
4.8.7. Comparing Effort Between Learners Grouped by Device Type (H31).....	202
4.8.8. Comparing Frustration Between Learners Grouped by Device Type (H32).....	204
4.8.9. Summary of Results for Comparisons between Device Type (Phone vs. Tablet).....	205
4.9. RESULTS RELATING TO SCREEN SIZE.....	206
4.9.1. Comparing Net Task Load Between Learners Grouped by Screen Size (H33)	208
4.9.2. Comparing Test-Scores Between Learners Grouped by Screen Size (H34).....	210
4.9.3. Comparing Mental Demand Between Learners Grouped by Screen Size (H35).....	211
4.9.4. Comparing Physical Demand Between Learners Grouped by Screen Size (H36).....	212
4.9.5. Comparing Temporal Demand Between Learners Grouped by Screen Size (H37)	213
4.9.6. Comparing Performance Between Learners Grouped by Screen Size (H38)	214
4.9.7. Comparing Effort Between Learners Grouped by Screen Size (H39)	215
4.9.8. Comparing Frustration Between Learners Grouped by Screen Size (H40)	217
4.9.9. Summary of Results for Comparisons between Learners Grouped by Screen-Size ...	219
4.10. RESULTS RELATING TO SPATIAL RESOLUTION OF DEVICE DISPLAY	220
4.10.1. Comparing Net Task load Between Learners Grouped by Screen Resolution (H41) .	220
4.10.2. Comparing Test-Scores Between Learners Grouped by Screen Resolution (H42)	222
4.10.3. Comparing Mental Demand Between Learners Grouped by Screen Resolution (H43)	223
4.10.4. Comparing Physical Demand Between Learners Grouped by Screen Resolution (H44)	225
4.10.5. Comparing Temporal Demand Between Learners Grouped by Screen Resolution (H45)	229
4.10.6. Comparing Performance Between Learners Grouped by Screen Resolution (H46)...	230
4.10.7. Comparing Effort Between Learners Grouped by Screen Resolution (H47)	231
4.10.8. Comparing Frustration Between Learners Grouped by Screen Resolution (H48)	232
4.10.9. Summary of Results for Comparisons between Learners Grouped by Screen Resolution.....	233
4.11. SUMMARY OF ALL STATISTICALLY SIGNIFICANT RESULTS.....	234
4.12. SUMMARY OF CHAPTER 4	237
5. CHAPTER 05 – ANALYSIS AND DISCUSSION.....	238
5.1. EVALUATION OF PARTICIPANT DEMOGRAPHICS	239

5.2. EVALUATION OF FINDINGS BETWEEN GROUPS	240
5.2.1. Evaluation of Results Comparing Net Task load Between Groups (H1)	240
5.2.2. Evaluation of Results Comparing Net Task load Between Smartphone Learners and the Control Group (H9-H16)	241
5.2.3. Evaluation of Results Comparing Task load between Tablet Learners and the Control Group (H17-H24)	248
5.2.4. Evaluation of Results Comparing Test Scores between Groups (H2)	249
5.2.5. Evaluation of Results Comparing Mental Demand Between Groups (H3)	250
5.2.6. Evaluation of Results Comparing Performance Between Groups (H6)	251
5.2.7. Evaluation of Results Comparing Physical Demand Between Groups (H4)	254
5.2.8. Evaluation of Results Comparing Temporal Demand Between Groups (H5)	254
5.2.9. Evaluation of Results Comparing Effort Between Groups (H7)	255
5.2.10. Evaluation of Results Comparing Frustration Between Groups (H8)	257
5.2.11. Summary of the Evaluation of Principal Findings	258
5.3. EVALUATION OF SECONDARY FINDINGS	259
5.4. EVALUATION OF RESULTS BETWEEN SMARTPHONE AND TABLET LEARNERS	260
5.4.1. Evaluation of Results Relating to Task load Between Smartphone and Tablet Learners (H25-H32)	260
5.5. EVALUATION OF RESULTS BETWEEN LEARNERS GROUPED BY SCREEN SIZE OF DEVICE USED	263
5.5.1. Evaluation of Results Relating to Net Task load vs. Screen Size (H33)	263
5.5.2. Evaluation of Results Relating to Effort and Screen Size (H39)	266
5.5.3. Evaluation of Results Relating to Frustration and Screen Size (H40)	266
5.6. EVALUATION OF RESULTS FOR COMPARISONS BETWEEN LEARNERS GROUPED BY SPATIAL RESOLUTION OF DEVICE DISPLAY	268
5.6.1. Evaluation of Results Relating to Differences between Retina-Display vs. Non-Retina-Display Devices	270
5.6.2. Evaluation of Results Relating to Correlations between Display Resolution In Pixels per Inch and the Subscales of the NASA TLX.	273
5.7. SUMMARY OF CHAPTER 05	276
6. CHAPTER 06 - CONCLUSION	279
6.1. RESEARCH PURPOSE AND INTENTIONS	279
6.2. RESEARCH BOUNDARIES	283
6.2.1. Learning Theory	283
6.2.2. Generalisability	283
6.2.3. Methodology	283
6.3. CONCLUSIONS	284
6.3.1. Answer to The Research Question	284
6.3.2. Secondary Conclusions	286
6.4. ORIGINAL CONTRIBUTION TO KNOWLEDGE	290
6.4.1. Theoretical Contributions	290
6.4.2. Methodological Contributions	291
6.4.3. Contributions to Professional Practice	292

6.5. CRITIQUE OF THE RESEARCH	292
6.5.1. <i>Validity</i>	292
6.5.2. <i>Reliability</i>	297
6.5.3. <i>Bias</i>	297
6.6. AGENDA FOR FURTHER RESEARCH	301
6.7. REFLECTION ON INTELLECTUAL AND PROFESSIONAL DEVELOPMENT.....	304
6.7.1. <i>Intellectual Development</i>	304
6.7.2. <i>Professional Development</i>	307
7. REFERENCES.....	309
8. APPENDICES.....	337

I. List of Figures

Figure 1-1: Theoretical framework for the study	14
Figure 1-2: A computational model of cognitive processing (after Putnam (1961), Carpenter and Reddi (2012) and Neisser (1967)	18
Figure 1-3: m-learning as a subset of CBL, note that the section-sizes are not intended to be representative of the scale of employment.	32
Figure 2-1: The FRAME model (Koole and Ally, 2006). The three circles of the diagram represent the device usability (A), the learner aspect (B) and the social aspect (C) of m-learning.	53
Figure 2-2: A conceptual map providing a technical and pedagogical overview of m-learning.	57
Figure 2-3: model of working memory proposed by Baddeley and Hitch (2000)	60
Figure 2-4: (after van Merriënboer and Sweller, 2010) ICL and ECL are additive, if they exceed the capacity of the working memory, germane resources cannot be employed in the formation of schemata.	63
Figure 2-5: (after Hollender et al., 2010; van Merriënboer and Sweller 2010) Proposed ICL and ECL in m-learning	99
Figure 3-1: Replica human skulls, plastic and resin-cast models.....	137
Figure 3-2: Screenshot of the application used in the mobile-device learning-task application (Samsung Galaxy S7 smartphone)	139
Figure 3-3: Control group learning materials, a labelled photograph.....	140
Figure 3-4: Flow chart showing the decision-making process used to determine appropriate statistical tests	155
Figure 4-1: Normal distribution of the weighted NASA TLX data from the experimental group.....	161
Figure 4-2: Normal distribution of the weighted NASA TLX data from the Control Group.....	161
Figure 4-3: Non-normal distribution of raw NASA TLX data in the experimental group	162
Figure 4-4: Boxplot showing outliers in the physical demand data (Y axis). See text for full explanation	163
Figure 4-5: (Normal) distribution of data in the net pre/post-test score for the experimental group	165
Figure 4-6: (Non-normal) distribution of data in the net pre/post-test score for the control group.....	166
Figure 4-7: Histogram showing comparison of the distribution of pre/post-test marks between groups	167
Figure 4-8: Pie chart showing experimental group gender ratio.....	172
Figure 4-9: Histogram showing experimental group age distribution and mean	172
Figure 4-10: Pie chart showing control group gender ratio	173
Figure 4-11: Histogram showing control group age distribution and mean	173
Figure 4-12: NASA TLX (mean) net task load score between groups.....	174
Figure 4-13: Boxplot of net test-score between groups (lenient marking scheme) (median values shown) ..	176
Figure 4-14: Boxplot of net test-score between groups (strict marking scheme) (median values shown)....	177
Figure 4-15: Boxplot of reported levels of mental demand between groups (median values shown)	178
Figure 4-16: Boxplot of reported levels of physical demand between groups (median values shown).....	179
Figure 4-17: Reported levels of temporal demand between groups	180
Figure 4-18: Boxplot of reported level of performance between groups (median values shown).....	182
Figure 4-19: Scatterplot showing weak negative correlation between performance and net test score.....	183
Figure 4-20: Boxplot of reported level of effort between groups (median values shown)	184
Figure 4-21: Scatterplot showing a medium positive correlation between mental demand and effort	185
Figure 4-22: Boxplot of reported level of frustration between groups (median values shown)	186
Figure 4-23: NASA TLX (mean) net task load score between non-interactive learners and smartphone-learners	188
Figure 4-24: Boxplot of reported levels of mental demand between groups (median values shown)	190
Figure 4-25: Boxplot of reported level of effort between groups (median values shown)	192
Figure 4-26: Boxplot of reported levels of physical demand between groups (median values shown).....	193
Figure 4-27: Pie chart showing distribution of device type	194
Figure 4-28: NASA TLX (mean) net task load score by device type.....	195
Figure 4-29 Pre/post-test scores by device type.....	197
Figure 4-30: Boxplot of reported mental demand by device type (median values shown)	198
Figure 4-31: Boxplot of reported physical demand by device type (median values shown).....	199

Figure 4-32: Temporal Demand by Device Type.....	200
Figure 4-33: Performance by device type.....	201
Figure 4-34: Effort by device type.....	202
Figure 4-35: Frustration by device type.....	204
Figure 4-36: Pie chart showing distribution of mobile device screen sizes in inches.....	206
Figure 4-37: Histogram showing distribution of mobile device screen sizes in inches	206
Figure 4-38 NASA TLX (mean) score by Device Screen Size	208
Figure 4-39: Pre/post test score by screen size	210
Figure 4-40: Boxplot of mental demand by screen-size (median values shown).....	211
Figure 4-41: Boxplot of physical demand by screen size (median values shown)	212
Figure 4-42: Boxplot of temporal demand by screen size (median values shown).....	213
Figure 4-43: Performance by device screen size.....	214
Figure 4-44: Boxplot of reported effort by device screen size (median scores shown).....	215
Figure 4-45: Pairwise comparisons of effort by device screen size.....	216
Figure 4-46: Reported frustration by device screen-size	218
Figure 4-47: NASA TLX (mean) score by spatial resolution of device display	220
Figure 4-48: Scatterplot showing weak negative correlation between NASA TLX net task load and device screen resolution	221
Figure 4-49: Scatterplot showing a weak negative correlation between mental demand and device screen resolution	224
Figure 4-50: Boxplot of Physical Demand by Display Resolution (median values shown).....	225
Figure 4-51: Scatterplot showing weak negative correlation between physical demand and device screen resolution	226
Figure 4-52: Scatterplot showing moderate positive correlation between physical demand and temporal demand in the non-retina-display group.....	227
Figure 4-53: Scatterplot showing no significant correlation between physical demand and temporal demand in the retina-display group	228
Figure 5-1: Diagram of The Cognitive Theory of Multimedia Learning (Mayer, 2005, p.37)	243
Figure 5-2: Spatial contiguity - The mobile app (right) permitted the labels to be spaced close to the structures on the photograph	245
Figure 5-3: Performance vs. task load - adapted from Chen, et. al. (2016, p.39).....	251
Figure 5-4: Scatterplot showing a weak correlation between performance and pre/post-test-score	253
Figure 5-5: Scatterplot showing a moderate positive correlation between mental demand and effort.....	255
Figure 5-6: Scatterplot showing a moderate correlation between effort and mental demand in the mobile-learners group	263
Figure 5-7: Device feature improvements by release date (Han and Cho, 2016)	264
Figure 5-8: Resolution comparison in halftone printing (magnified)	268
Figure 5-9: Comparative spatial screen-resolution (magnified) of 400 PPI (left) and 100 PPI (right)	269
Figure 5-10: Scatterplot showing moderate correlation between physical and temporal demand in non-retina-display device learners	271
Figure 5-11: Scatterplot showing moderate negative correlation between screen size and spatial resolution	272
Figure 5-12: Scatterplot showing weak negative correlation between spatial resolution and net task load	273
Figure 5-13: Scatterplot showing weak negative correlation between physical demand and spatial resolution	274
Figure 5-14: Scatterplot showing weak negative correlation between spatial resolution and mental demand	275

II. List of Tables

<i>Table 3-1 Effect size thresholds used for parametric tests (Cohen 1988).....</i>	<i>153</i>
<i>Table 3-2 Effect size thresholds for non-parametric tests.....</i>	<i>154</i>
<i>Table 4-1 Group Statistics for NASA TLX Score between groups.....</i>	<i>175</i>
<i>Table 4-2 Group statistics for NASA TLX score between groups.....</i>	<i>181</i>
<i>Table 4-3: Summary of all comparisons between experimental group and control group.....</i>	<i>187</i>
<i>Table 4-4 Group Statistics for NASA TLX Score between groups.....</i>	<i>189</i>
<i>Table 4-5 Group Statistics for NASA TLX Score by Device Type.....</i>	<i>195</i>
<i>Table 4-6: Group Statistics for effort by Device Type.....</i>	<i>202</i>
<i>Table 4-7: Summary of all comparisons relating to device type.....</i>	<i>205</i>
<i>Table 4-8: Statistically significant pairwise comparisons of effort by device screen size.....</i>	<i>216</i>
<i>Table 4-9: Multiple comparisons of frustration by screen-size (Tukey HSD).....</i>	<i>218</i>
<i>Table 4-10: Summary of all results for comparisons between learners grouped by device screen size.....</i>	<i>219</i>
<i>Table 4-11: Summary of results between learners grouped by spatial resolution of device display.....</i>	<i>233</i>
<i>Table 4-12: Statistically significant results between mobile-learners and non-interactive learners.....</i>	<i>234</i>
<i>Table 4-13: Statistically significant results between smartphone-learners and non-interactive learners.....</i>	<i>234</i>
<i>Table 4-14: Statistically significant results between tablet-learners and non-interactive learners.....</i>	<i>234</i>
<i>Table 4-15: Statistically significant results relating to tests between mobile-learners grouped by device type.....</i>	<i>235</i>
<i>Table 4-16: Statistically significant results between m-learners grouped by size of device display.....</i>	<i>235</i>
<i>Table 4-17: Statistically significant results between m-learners grouped by resolution of device display.....</i>	<i>235</i>

III. List of Appendices

APPENDIX A :	THE ONLINE DATA COLLECTION METHOD (DISCONTINUED).....	338
APPENDIX B :	THE NASA TLX DATA COLLECTION FORM.....	356
APPENDIX C :	PRE/POST-TEST DATA COLLECTION FORM.....	358
APPENDIX D :	PARTICIPANT INFORMATION AND CONSENT FORM	359
APPENDIX E :	ETHICAL APPROVAL.....	363
APPENDIX F :	HARDWARE AND SOFTWARE RESOURCES.....	365

COPYRIGHT DECLARATION

Attention is drawn to the fact that copyright of this thesis rests with:

- I. Anglia Ruskin University for one year and thereafter with
- II. John Stephen Talbot

A copy of this thesis has been supplied on condition that anyone who consults it is bound by copyright.

1. CHAPTER 01: INTRODUCTION

Smartphones such as the Samsung Galaxy series (Samsung Electronics.) and the Apple iPhone (Apple Inc.) are a global phenomenon. At the time of writing, there are currently over 2.6 billion smartphones and tablet computers in use, accounting for 87% of the worldwide market share for connected devices (Shirer, 2015). Their ubiquity and small size make smartphones an ideal medium for the delivery of distance learning materials, leading to a burgeoning new field of education known as mobile learning (m-learning) (Pachler, Bachmair and Cook, 2010). The major publishing houses have been quick to spot the potential of such devices as a replacement or adjunct to paper editions of textbooks, offering most popular editions as electronic versions that can be downloaded to smartphones or tablet devices such as the Kindle (Amazon Inc.). Literature is a key information source in education, and for distance-learners, books have long been the only educational materials that have the portability factor required for any-time, any-place learning. Smartphones and electronic publications are set to challenge this position. Like books, their pocketable size allows them to be used in remote locations other than the classroom, but unlike books the technology afforded by such devices also permits the inclusion of interactive multimedia content such as audio, video and graphic animation (Mayer, 2009). Modern smartphones are sophisticated computers, and the use of computing devices and their associated software is known to impact on the cognitive resources of the operator. Hardware and software design is informed by a field known as Human Computer Interaction (HCI). Hurtienne (2009, p.12) states that HCI is a discipline that is “deeply rooted in cognitive science” and is used to optimise the design of computing devices and interfaces to reduce cognitive load in the user. Touch-screen devices such as smartphones present a new challenge to HCI due to their technological complexity and their departure from the usual mouse-driven graphical interface. New approaches have been devised, known as “third wave” HCI to inform the design of smartphone interaction and increase user-friendliness by reducing cognitive load. In education, cognitive load is also a central concept. Cognitivist and constructivist educational theory both require that the cognitive faculties of the learner are optimised. This understanding has led to the development of Cognitive Load Theory (CLT) (Sweller,

1989) which is used to ensure that learning materials are designed in a way that does not overload the cognitive capacity of the learner. Given that computers such as smartphones are known to contribute to the cognitive load of the user, and that these devices are now being used to deliver learning content such as electronic books, it is proposed that there is a need to investigate whether smartphones offer any net cognitive advantage to the learner when used in lieu of traditional textbooks.

In summary, this thesis examines the new phenomenon of touch-screen mobile computing devices such as smartphones and tablet computers and seeks to justify their use in the field of learning from a cognitivist perspective. CLT (Sweller, 1989), the Baddeley and Hitch Model of Working Memory (1974) and the Cognitive Theory of Multimedia (Mayer, 2009) will be used to provide a conceptual framework for the study, which will also incorporate the principles of HCI (Longuet-Higgins, 1981; Card, Newell and Moran, 1983) in looking at the design of software and hardware for the delivery of learning materials. Cognitive load and other factors affecting task load will be measured in learners using mobile touchscreen devices and compared to learners studying a labelled textbook photograph to assess whether there are any significant differences and whether these differences have any impact on learning.

1.1. RESEARCH AIM

The primary aim of the study is to quantify whether the use of a mobile touch-screen device (a smartphone or tablet) as a learning tool, has any adverse or beneficial effect on the learning task load imposed upon a learner in comparison to a traditional learning activity, such as studying a labelled photograph in a textbook. For the purposes of this study, smartphones and tablets are defined as “a form of mobile personal computer with a large, touch-sensitive screen operated using a pen, stylus or finger” (Atkinson, 2008, p3).

The secondary aim of the study is to critically evaluate the effect of other independent variables on task load. These include the screen-size of the mobile device and the spatial resolution of the device display. These variables will be assessed to determine any effects that they may have on mental and physical demand, time-pressure, performance, effort and user frustration.

1.2. RESEARCH OBJECTIVES

The objectives of the study are:

- To determine the net task load placed on mobile-learners by an interactive learning activity presented on a mobile device and compare this with a control group of non-interactive learners to assess whether there is any statistically significant difference.
- To determine a pre/post-test score for mobile-learners to assess whether an m-learning activity resulted in successful learning of the structures of the human skull base, and compare this with a control group of non-interactive learners to assess whether there is any statistically significant difference.
- To determine task load-scores relating to contributing sub-scales of mental demand, physical demand, temporal demand, performance, effort and frustration placed on mobile-learners by an interactive learning activity presented on a mobile device and compare this with a control group of non-interactive learners to assess whether there is any statistically significant difference.
- To determine task load-scores relating to mental demand, physical demand, temporal demand, performance, effort and frustration placed on mobile-learners by an interactive learning activity when grouped by device type; namely smartphone vs tablet computer.
- To determine task load scores relating to mental demand, physical demand, temporal demand, performance, effort and frustration placed on mobile-learners by an interactive learning activity when grouped by device screen-size.
- To determine task load-scores relating to mental demand, physical demand, temporal demand, performance, effort and frustration placed on mobile-learners by an interactive learning activity when grouped by the spatial resolution of the device display.

1.3. A NOTE ON WRITING STYLE

The reflective nature of an educational doctorate demands a first-person perspective when writing. However, the experimental nature of the methodology used in this study requires a scientific approach that typically favours a distancing of the researcher's voice in the narrative. For this reason, I have attempted to reconcile these differences by using a passive voice in the non-reflective sections of the thesis, and a first person writing style in the reflective sections. To avoid confusion, I have consistently used the term "the author" to refer to other researchers whose papers are cited rather than to myself.

1.4. PROFESSIONAL BACKGROUND

A professional doctorate, by definition, requires reflection on professional practice. Sociologist Anthony Giddens cites reflexivity as an essential tool in the field of life-long-learning (Hall,2002). As society changes, so the self must adapt. In Giddens' terms, one becomes a *reflexive project* and it only by this self-exploration that one can keep pace with change. Hall (p.3) acknowledges the on-going nature of the process. He sums up Giddens' reflexive construction of self-identity by saying that "self-reflexivity entails the seeing of one's life as a project always in the making". In short, reflexive practice allows the location and adaptation of the self within the world in which one operates and fosters the approach of life-long learning. The construction of this self-knowledge will also be affected by cultural and traditional values (i.e. behaviour that we think is expected of us by others), ethics and morality. Locating the position of one's self in the research is important for several reasons:

Firstly, it is likely that the choice of topic, research question and methodology will relate to one's professional practice setting.

Secondly, the ethically-centred principles of respect, dignity, confidentiality and *primum non nocere* - first do no harm - apply to both research participants and researchers.

Thirdly, from the viewpoint of experimental bias in empirical studies, it is usually desirable for the researcher to have a detachment from what is being observed or measured (Rugg and Petre, 2007).

Finally, it is likely that the very act of research will stimulate one's own development intellectually and professionally and that any conclusions drawn from the research will have implications for future practice. There are three aspects to my professional practice that led me to consider this research topic. These are:

- previous clinical practice
- my current role as a computer-based learning facilitator
- my scholarly activity as a writer

In the following sections, my aim is to provide a short overview of each of these professional facets as they relate to the theme of this study. I will attempt to define my

position in the field of education research, justify my professional reasons for undertaking research and examine some of the potential issues relating to research within my professional practice area of m-learning.

1.4.1. Professional Practice: Clinical Experience

Before taking up my lectureship, my background was in computerised medical imaging. My interests lay principally in the field of neuroscience, as a research radiographer at both the Walton Centre for Neurology and Neurosurgery and Oxford University. This health-research background, coupled with my present role as a distance-learning lecturer and software designer largely dictates my philosophical stance regarding research. The studies in which I have previously been involved tended to use an experimental approach common to the physical sciences. The two underpinning themes behind this thesis are technology and cognition, both relate strongly to physical science, and both disciplines have been identified as having much potential in education (Laurillard, 2008a; Traxler, 2009; Kukulska-Hulme et al., 2010). When looking at how learning is mediated via technology, there are several key themes that emerge from the literature that are closely related to my professional practice particularly:

- Creating content specifically designed for mobile devices
- Setting up new educational practices supported by mobile technology
- Designing tools and infrastructures to make content available on new devices
- Instructional design as applied to e-Learning.
- Cognitive redundancy and split-attention effects due to the potential for complexity in e-learning.

(Asraj, Freeman and Chandler, 2011; Arrigo et al., 2013)

My professional practice blends clinical knowledge as an educator in the field of medical imaging, with the technical and pedagogical skills required to facilitate distance learning. In 1989, I was introduced to the concept of Nuclear Magnetic Resonance Imaging, now known as Magnetic Resonance Imaging or MRI. I have specialised in this subject, both as a practitioner, and over the last 25 years as an educator. On reflection, my interest in

education, and human cognition both stem from working in this field. The topics relating to the practice of MRI are not taught at undergraduate level, so between 1995 and 1997 I undertook a Master's degree in clinical MRI. On completion of this degree, I was employed as a research radiographer by Oxford University, and through this post, I became involved with various educational activities. These included a short course in MRI that I helped to develop and formal collaborations with Anglia Ruskin University. My involvement with MRI and particularly neurological MRI has had a significant influence on my pedagogical practice, not just from the viewpoint of writing and illustrating books and course materials, but also from physiological and psychological perspectives. Over the last 15 years, functional MRI of the brain (fMRI) has been seeking to identify the neurological processes involved in learning and the working memory. (Delazer, et al., 2003, Manoach, et al., 1997) Despite recent criticisms by Eklund, Nichols and Knutsson, (2016) fMRI is capable of showing that learning involves physiological processes and the areas of the brain cortex involved in learning can be identified. For example, when a subject is required to learn something by both seeing and hearing information, different areas of the cortex are simultaneously involved. This neuroscientific background first caused me to consider how learning could be quantified from an educational-psychology perspective.

1.4.2. Professional Practice: Computer-Based Learning

My Interest in computer-based learning and the use of multimedia stems from an opportunity that was provided to me in 1999. I was approached by the lead radiologist from Europe's largest clinical MRI course to create interactive digital learning resources for the course delegates. The concept was an early example of a *flipped-classroom* approach whereby the factual lectures were provided to the delegates on compact-disc (CD) in advance of the conference. This pre-learning allowed more time to be devoted to group discussion during the live sessions. At the time, the technology required for this type of computer-based learning (CBL) was still relatively primitive. Using a combination of digital media and hypertext mark-up language (HTML) I created a CD-based interactive interface that allowed users to watch and control digitised versions of live lecture presentations. Over the following three years, I developed this tool, taking advantage of the new audio and video compression algorithms that facilitated online access to streaming downloadable digital media such as mp3 (Motion-Picture Experts Group, Audio Layer III). Users could navigate the lectures in a non-linear way. When two presenters covered the same topic, users could jump from one presentation to another. The benefits of this mode of delivery were that learners could control the pace of the presentation, instantly revisit sections that they did not understand, replay entire lectures at a later date or browse by topic or by speaker. By the fourth edition of the CD, I used third-party authoring software (Director, Macromedia Inc.) that was able to compress over 12 hours of clinical lectures into a file of just 500MB in size. Individually, the compressed lectures were small enough to be watched in an internet browser, even without broadband connectivity. This type of functionality has now become commonplace, but I feel that there are new opportunities for digital learning as technology progresses into a new era of touch-screen devices and mobile applications. These devices not only allow tactile control over multimedia materials but their compact form also permits the user to engage with learning in physical locations other than at home or in the classroom.

In 1997, the Dearing report devoted an entire chapter (13) to the role of Information technology in higher education and stated that:

“...we believe that the innovative exploitation of Communications and Information Technology (C&IT) holds out much promise for improving the quality, flexibility and effectiveness of higher education. The potential benefits will extend to, and affect the practice of, learning and teaching and research... There is scope to reduce costs in the future and the potential is great, but implementation requires investment in terms of time, thought and resources in the short term.” (Dearing, 1997, p.202)

This paragraph makes two key points, firstly that computers offer numerous benefits to education, but also that implementation of digital learning materials requires considerable resources and development time. It transpired that this was not just in the short-term. The continuing task for distance learning educators is to develop programmes that effectively harness new technology (such as mobile devices) while fostering a “deep learning” approach, founded on established and proven teaching methods (White, 2000). I think there have been over-expectations from the academic community about what e-learning can offer, and how quickly we can understand and utilise its full capabilities. Many of the short-term issues associated with the early implementation of e-learning have now been addressed. These obstacles related to factors such as slow network speed (Keller, 2011), the lack of user-friendly authorware, educator resistance (Kopcha, 2012) and the low processing power of earlier computing devices. However, a new challenge has arisen. With advances in IT technology, hand-held mobile devices have become a primary source of information. This new technology means that computer-based learning materials are no longer restricted to static, classroom-based machines. Instead, they offer location-independent access to media-rich, visual content with video, audio and interactive elements, music, and animated graphics (Mayer, 2012). Importantly, because of their improved computing power, portability and long battery-life, modern mobile devices have brought a flexibility to where computer-based learning can occur that is greater than anything that has been available before. Formerly non-productive periods of time, away from traditional learning materials, can now be gainfully employed in learning.

Freedom to learn independently of location is relevant to most of my students, who are engaged in the wider community of service-learning, extending their experience beyond that which is possible through a purely classroom-based context. Much of the learning

takes place in a clinical setting, and no attendance on campus is required. Computer-based learning from a mobile device such as a tablet PC or smartphone gives these students a very high degree of freedom in when, where and how they can learn. This new medium has led me to create interactive digital versions of all the lectures that are delivered to classroom-based students, in a format that can be viewed in mobile applications (apps). Some of the topics relevant to these learners are well suited to the digital format. The operation of a computer-based medical-imaging modality such as MRI can be simulated to a high degree on personal and mobile computing devices, and the physics of MRI involves molecular motion and vector models that can be replicated using computer generated imagery.

For the reasons above I became interested in touch-screen devices as learning and teaching tools and noted the arrival of the iPad (Apple Inc.) in 2010 with interest. This new type of touch-screen interface caused me to think about what types of subject would be most appropriate to these devices and what types of learner would particularly benefit from their use. With these questions in mind, I registered as a mobile application developer in 2011 and have had some success in creating (commercially available) distance-learning materials for iPhone, iPad and Android devices (running the Google operating system).

One of the main topic-areas that I identified as being suited to the touch-screen device is human anatomy. Teaching anatomy at a distance presents challenges. Traditionally, anatomy is taught in the dissection room. By its nature, this method of teaching offers a high degree of authenticity, but interestingly, in a study of 2007, anatomy students gained higher marks and a better pass-rate when a blended-learning approach was used (Pereira, et al., 2007). The blended-learning students had access to animated multimedia, interactive quizzes and communication via online fora as well as downloadable printed materials. A more recent study by Lewis, et al. (2014) showed similar results when students used 3D anatomy applications to enhance laboratory learning. These apparent benefits of interactive multimedia in the teaching of anatomy led me to reflect on whether touch-screen devices could be used to replicate the authenticity of the anatomy lab, and in doing so offer a stand-alone m-learning solution to the teaching of anatomy. Currently, the processing power and graphics architecture of mobile devices permit high

definition computer-generated imagery that can represent physical objects very accurately, and the touch screen allows those objects to be manipulated in virtual space. Pinching and swiping gestures can enlarge and rotate the virtual objects in much the same way as an anatomical specimen can be manipulated in the physical world. The devices, therefore, appear to offer a high degree of interaction and authenticity than cannot be achieved using textbooks or electronic documents alone. This theme is more fully-covered in the literature review on page 52.

Anatomy is well-suited to m-learning from a pedagogical perspective. The remote nature of distance learning does not easily lend itself to a constructivist approach in which learners build knowledge through discourse and group problem-solving (Vygotsky, 1986; Piaget, 2001). The remote learner is typically isolated from a group, autonomous and is restricted somewhat to a more didactic lecture (or materials-driven) strategy whereby instructional content such as ePresentations are served by a computer. Learning anatomy often requires the learner to memorise structures, such as the names of bones, nerves and blood vessels. This type of rote-learning, although sometimes considered to be superficial by educationalists, is essential when studying human anatomy. It is also ideally-suited to m-learning where learners are typically remote from other students, precluding the use of group activities. Although this method of learning appears to fly in the face of contemporary educational thinking, the superiority of a cognitivist approach over constructivism has also been demonstrated in computer-based learning by Vogel-Walcutt, et al. (2011). The authors discovered that the use of simulation-based training incorporating CLT lead to better retainment of integrated knowledge in higher levels of the cognitive domain. In summary, m-learning seems ideally suited to learning anatomy.

1.4.3. Professional Practice: Scholarly Activity

As a lecturer, I am expected to undertake a degree of research and scholarly activity in addition to teaching. A large part of this involves writing textbooks for publication. During my employment as research radiographer at University College Oxford, I was invited to co-author and provide scientific and anatomical illustrations for three popular books in the field, *MRI in Practice*, *MRI at a Glance* and *Handbook of MRI Technique* (Wiley-Blackwell).

Despite the value of printed materials, from a professional perspective, one of the lecturer's principal roles is to engage the learners and encourage their interaction with new ideas and new learning materials – including new technology (Gagné, 1985, 1988; Mayer, 2009). Mobile devices offer a similar portability-factor to books and in some cases are less expensive to purchase. The possibility offered by mobile devices to access materials, communicate and write notes has considerable potential. My publisher (Wiley-Blackwell) is now considering whether future electronic editions of my books should include interactive elements featuring computer generated imagery. A companion website already offers some of these features. However, my initial literature review indicated that further research into this area is required to help make informed decisions about how the material is presented, to discover potential barriers to educator and learner-adoption, and discover to what types of learners it would be most applicable. (Hollender, et al., 2010; Liu, Li and Carlsson, 2010; Schmidt-Weigland, 2011; Aslanian and Clinefelter, 2012; Raptis, et al., 2013; Sung and Mayer, 2013).

Assessing strategies for distance learning is critical because these materials have to stand up on their own, without human delivery. To be fit for purpose they must fulfil the factors that underpin successful learning. As an author, and application developer, the possibility of using mobile devices to replace the use of textbooks in learning is a very pertinent area for research. It would be useful in devising the strategies employed for the creation of learning materials and the production of commercially available e-books - particularly in justification of the long production time and associated high cost of developing such materials (Chapman, 2006).

Mobile technology is still evolving rapidly, and the inherent complexity of multimedia content and device interfaces has the potential to adversely affect the ability of the user

to learn. One recurring theme in the topic literature relates to the way in which mobile devices impact on the cognitive processes of the learner. (Hollender, et al., 2010; NMC Horizon Report, 2012; Cheong, Bruno and Cheong, 2012; Arrigo, et al., 2013). An investigation into potential barriers to learning relating to mobile-device use would, therefore, be relevant and transferrable to practice when designing future learning materials and targeting learners. If m-learning offers no tangible benefit versus traditional text or hypertext, the investment of time and expense in developing for this platform may not be justified. If the opposite holds true and learning is enhanced, then the investment will be paid back many times over, as these materials are easily recyclable and reusable (Mayer, 2012).

1.5. CONCEPTUAL AND THEORETICAL FRAMEWORKS

1.5.1. Educational Theory and m-learning

In this section, I aim to present the theoretical framework that underpins my choice of research topic, and methodological approach. Maxwell (2012) explains that a conceptual framework is a graphic or narrative model showing key factors, concepts or variables to be studied. Figure 1-1 is intended to represent the overarching theoretical framework relating to my study. The key point that the diagram is intended to illustrate is the evolution of two parallel theories that are used to inform computer-based learning, namely Cognitive Load Theory and Human Computer Interaction Theory.

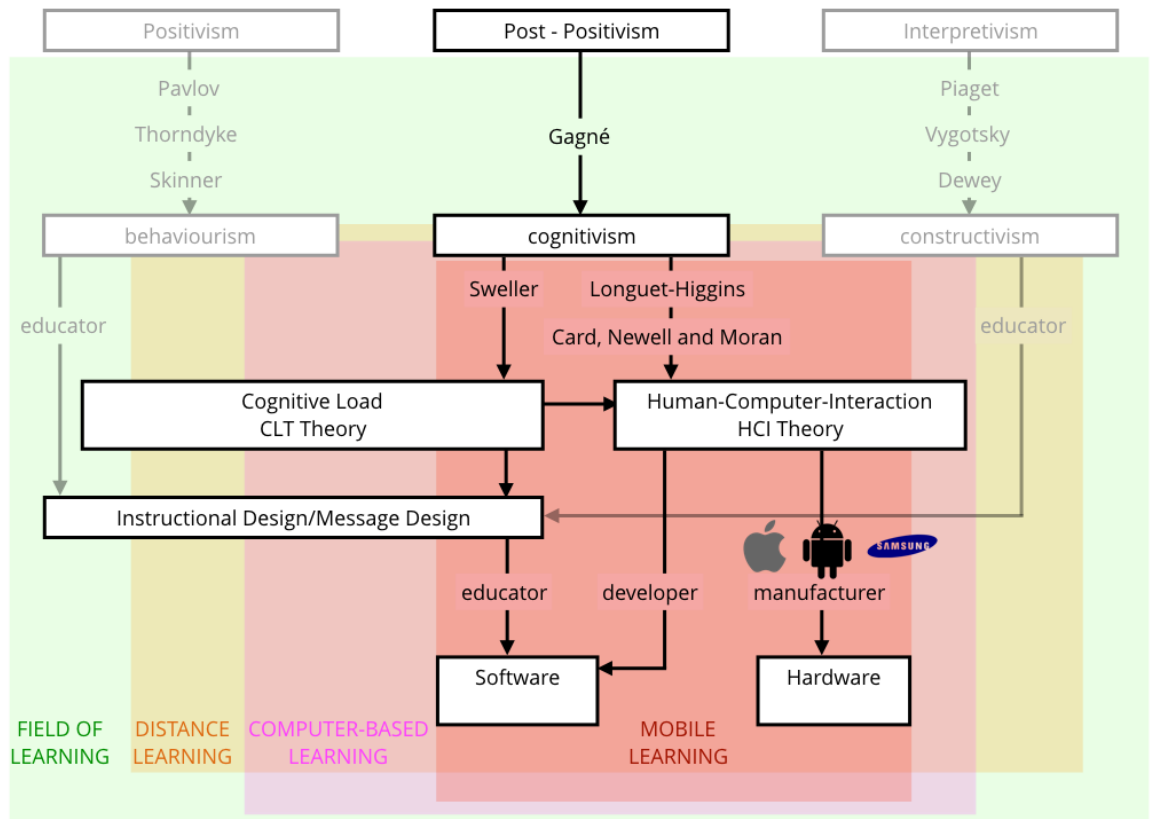


Figure 1-1: Theoretical framework for the study

The colour-shaded areas in Figure 1-1 represent subsections within the field of learning and identify that m-learning is a subset of computer-based learning that can be applied in

the classroom, or at a distance. The centre strand of the overlaid diagram (highlighted in bold) recognises that m-learning requires both hardware and software, the design of which is informed by:

- **Human/Computer Interaction** which can be used to inform software and device design (Zhang and Galletta, 2015). HCI develops the work of Gagné and has been adapted by researchers such as Longuet-Higgins (1981) and Card Newell and Moran (1983) to relate specifically to the human-computer interface. HCI is also of relevance to the distance learning educator as it ensures that software is accessible, intuitively easy to use and responsive.
- **Cognitive Load Theory** which can be used to inform instructional design used in educational (or commercial) software development (Sweller, 1989; Mayer, 2012). This theory is of relevance to a distance learning educator, as it seeks to ensure that learning materials are constructed in a way that does not create intrinsic barriers to learning.

These tandem theories are both highly relevant to m-learning and are evaluated in section 1.5.3. To the left of the diagram is a box representing instructional design and message design, this also relates to the design of the learning materials and takes into account:

- **The Cognitive Theory of Multimedia Learning**, a theory which combines elements of Sweller's Cognitive Load Theory (1989) and Paivio's Dual Coding Theory (1990). The underpinning concept being that "learners can better understand an explanation when it is presented in words and pictures, than when it is presented in words alone" (Mayer, 2009, p.3). This theory considers the mode of delivery (words, images), the media used (computer, smartphone) and the sensory modalities involved (auditory or visual).

1.5.2. Mobile Learning Through a Cognitivist lens

M-learning has its roots in the instructional design methodologies proposed by the likes of Benjamin Bloom and Robert Gagné. Over the first half of the 20th century, educational practice was informed by the behaviourist school of psychology. Behaviourism has its origins in the animal research of Russian physician Ivan Pavlov (1927) and is a perspective found in both psychology and in learning and teaching. The educational theory of behaviourism holds that the learner is a passive recipient of knowledge. In philosophical tradition, a metaphor of a blank slate or *tabula rasa* (erased wax tablet) is often used to represent the fact that the learner has no pre-existing knowledge - or more precisely lacks a particular desirable behaviour that the teacher wishes to instil (Aristotle, 2008). Behaviourism takes a positivist stance and was advocated by early practitioners such as J.B. Watson (1931) and B.F. Skinner (1988) as a backlash to introspectionism. Their rationale was based on the premise that data derived from self-analysis of mental processes could not be empirically measured, whereas changes in behaviour could be objectively observed and quantified.

The behaviourist approach was prevalent up until the end of the 1950s but the following decade saw a rejection of behaviourism in favour of an emerging theory known as cognitivism. The so-called *cognitivist revolution* of the 1950s (Baars, 1986) was driven by the emerging understanding that the human learning process involves more than reflex responses to external stimuli. In the pre-Socratic Greek philosophical origins of learning, there was a theory, in what is known as the rationalist tradition, that knowledge is *a priori*, meaning that humans have an innate knowledge that exists *before the event* of learning (Plato, 1973). This theory assumes that knowledge pre-exists in the mind of the learner waiting to be illuminated by the process of education. Behaviourist methods of classic conditioning, (and to a lesser degree operant conditioning) required the formation of mental links between innate responses and the required behaviour to be learnt (Skinner, 1988). However, modern neuro-scientific methods such as functional magnetic resonance imaging (fMRI) can provide new information about how learning occurs. Neuroscience tends to support a cognitivist *a posteriori* model where knowledge is not considered to be innate, and learning requires sensory input. New information is stored in

short-term and long-term memory *after the event* of learning (Moscovitch 1992, 1994; Mather, Cacioppo and Kanwisher, 2013).

The area of the brain that controls autonomic responses such as salivation (as researched by Pavlov) is known as the hypothalamus, but this is not the same area of the brain that is required when learning (for example) how to play a musical instrument or how to talk.

The hypothalamus is part of a structure known as the diencephalon. Its anatomical position is deep inside the brain, and it performs all of the autonomic functions required for an organism to survive, but the higher cognitive and motor functions required to learn a language, or how to play a musical instrument involve other parts of the brain, especially the outer cortex and sub-cortex (Carpenter and Reddi, 2012). There is now convincing evidence from neuro-imaging studies, that learning results in changes to brain plasticity, i.e. “changes in structure and function of the brain that affect behaviour and are related to experience or training” (Herholz and Zatorre, 2012, p.486) suggesting that knowledge does not pre-exist, but is (physiologically) constructed during the learning process. This concept has recently been confirmed in animal studies by Ramirez, et al., (2013) who were able to isolate memories relating to learning (external fear stimulus) in a small number of individual neurones (cells) in the brain. These studies convincingly discredit the Cartesian, rationalist view of *a priori* knowledge.

Cognitivism, therefore, considers the brain (metaphorically) as a computer requiring an input of data that must then be processed, stored in memory and then used to inform future actions. There is also a parallel strand of cognitivism known as computationalism, a theory, first established by philosopher Hilary Putnam in 1961. Computationalism acknowledges the similarities between a human brain and a computer. These include the inputting of information, the step-by-step algorithms required to process that information in random access memory (RAM), and importantly the ability to correlate one set of data with another.

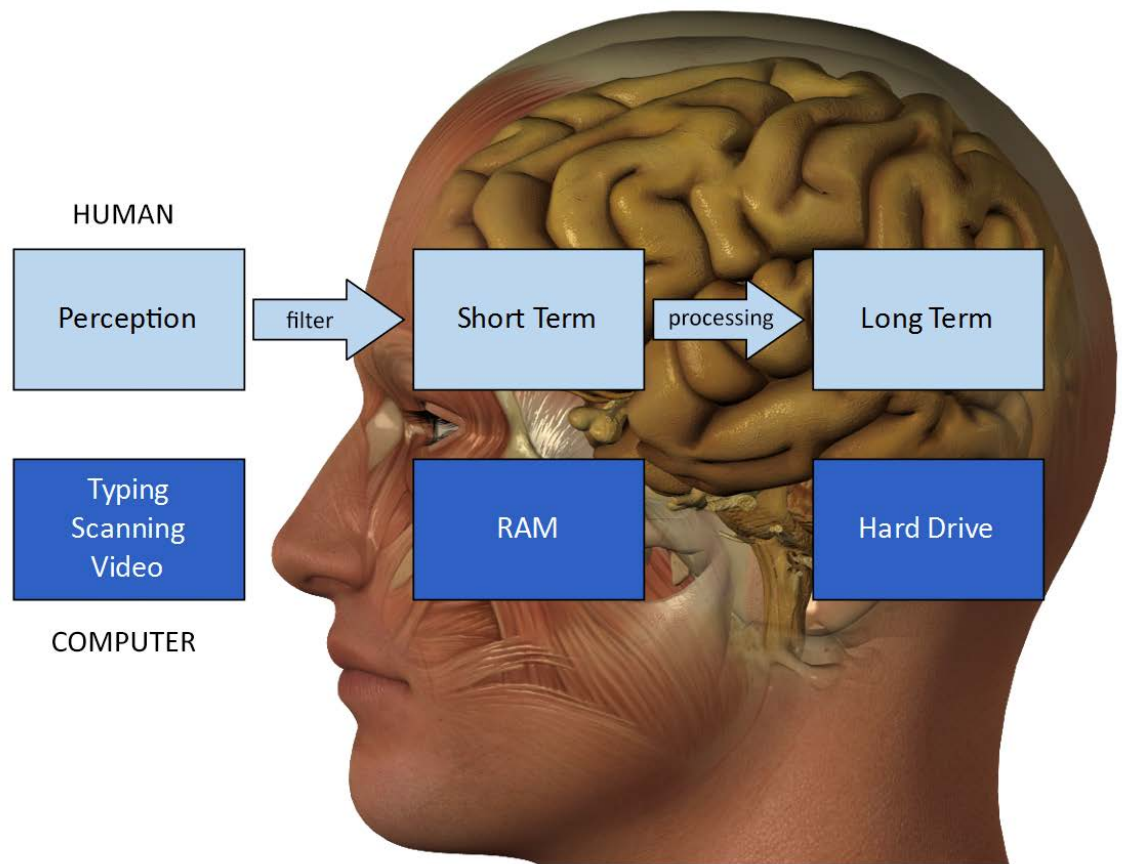


Figure 1-2: A computational model of cognitive processing (after Putnam (1961), Carpenter and Reddi (2012) and Neisser (1967))

The theory was further developed by Ulric Neisser (1967), the founder of cognitive psychology and neuroscientist David Courtney-Marr. Being able to correlate one set of data with another is crucial to learning the higher-level skills in Bloom's cognitive domain. Neuroscientists, such as Ramirez (2013), call these organised patterns *memory engrams*, but psychologists describe them as *schemata* (singular *schema*). This term was introduced by cognitivist educator and child-development researcher Jean Piaget (2001) and is key to cognitive learning. Piaget was one of the first educators to recognise the importance of Cognitivism in education. His interest was piqued during research into child development during which Piaget noted that some children gave strikingly illogical answers to simple questions. His subsequent research led to a theory of cognitive development that recognised that the biological development of the brain plays an important role in learning and that learning strategy must take cognitive development into account.

Piaget organised learning into stages, namely:

- Sensori-Motor (birth-2 yrs.)
- Pre-Operational (2-7 yrs.)
- Concrete Operational (7-11 yrs.) and
- Formal Operational (11 yrs. onwards)

Piaget's view differed from those held by many of his contemporaries, in that he proposed that these stages of cognitive constructivism are related to physiological maturity. This is certainly the case when considering the myelination of the brain in very young infants (Paus, 2005). Piaget also placed importance on learning as being created by personal experience and interaction with the environment rather than being reliant upon socio-cultural interactions and language – the theory favoured by Vygotsky (2012). Modern thinking has refuted some of Piaget's ideas. Bruner (1960) for example, argued that individuals have the capacity to learn fairly complex concepts at any stage in their development and that matching the level of learning activities to biological maturity may be of lesser consequence than Piaget believed. These differences aside, Piaget's schemata theory is still widely regarded in the field of cognitive educational-psychology. It provides a useful representation of mental activity that can be used in the design of teaching materials and assessing the cognitive load imposed on the learner. It also allows knowledge to be quantified because schemata are considered to be "chunks" of information. Carpenter and Reddi (2012) concur with Ramirez et al. (2013), in acknowledging that this process is mediated by the neurones of the brain and that this involves converting patterns of stimulation into patterns of response. This process relies on the fact that the human brain is structured in layers, each layer consisting of a network of neurones. Each neurone acts as a miniature computer in that it can respond to particular patterns of activity occurring the adjacent layers. The incoming sensory pattern generated by a stimulus is therefore modulated as the activity passes through the layers resulting in a very different pattern at the output (response). This process is echoed in the neural networks found in computing and is thought to be the underlying mechanism behind cognition.

In distance learning, one of the main challenges for an educator is the engagement of the learner at a location that is remote from the classroom. This physical separation is a potential pitfall in learning and teaching because cognitivist learning theory recognises the need for learner engagement from the outset of any learning activity. To foster engagement in the mobile learner, educational materials must be presented in a way that can be easily understood and assimilated (Sweller, 1994; Mayer, 2009; van Merriënboer and Sweller, 2010). This basic need relates strongly to how the human brain receives and processes information and therefore falls into alignment with what is known as the cognitive domain of learning. This domain was first identified by educational psychologist Benjamin Bloom (1956), whose interest centred around the design of educational materials, particularly relating to the setting of learning objectives. In Bloom's philosophy, these objectives should feature measurable outcomes relating to any learning that could be used to inform the student assessment process. With this in mind, Bloom spent several years from 1948 onwards, collaborating with many of the university examiners across North America. In 1956 this collaboration culminated in the publication of his influential work entitled *The Taxonomy of Educational Objectives, The Classification of Educational Goals, Handbook 1, The Cognitive Domain*.

The cognitive domain relates to brain-based learning and lies somewhat in opposition to the rationalist and behaviourist approaches that had been favoured for the first half of the 20th century. One of the first educators to recognise the importance of this trend towards a cognitive approach in learning was experimental psychologist Robert Gagné (1985), whose early work was very much in the behaviourist tradition. At the end of the 1950s - during the cognitivist revolution - Gagné switched his attention to the newly developing field of cognitivism. One of his principal interests at the time was instructional design. Using some of Bloom's themes relating to learning taxonomy and incorporating some of the concepts introduced by behaviourist B.F. Skinner, Gagné published an influential work in 1965 entitled *The Conditions of Learning*. He categorised learning into five main areas, namely;

Motor Skills relate to the ability to make co-ordinated body movements.

Verbal Information relates to the ability to state facts or describe something.

Attitude relates to “acquired internal state that influences the choice of personal action” (Gagné & Driscoll, 1988, p.58).

Intellectual skills relate to the mental faculty of reasoning.

Cognitive strategy relates to how “learners regulate their own internal processes of attending, learning, remembering, and thinking” (Gagné, 1985, p. 55).

Areas 4-5 relate to the way that the brain receives and stores information, particularly relating to verbal, intellectual and cognitive skills. These categories are pertinent to m-learning because they are reflected in the theory and process of designing interfaces required for human/computer interaction and software. Zhang and Galletta (2006) explain the need for HCI to be human-centred and list the human characteristics that are relevant to the interaction with information technology. It can be seen that they echo Gagné’s conditions of learning very closely, they are defined as:

Physical or Motor Skills relating to interaction with the user-interface.

Affective and Motivational Aspects relating to affective state, mood, feelings, emotions and motivation.

Cognitive Issues relating to perception, attention, memory, knowledge, learning, error and distributed cognition.

Demographics relating to gender age and culture.

The similarity with Gagné’s educational model is not coincidental, because HCI research evolved in tandem with the cognitivist revolution, and used many of Gagné’s underpinning theories relating to cognition in its methods. A cognitivist approach is, therefore, particularly suited to the development of m-learning materials and hardware design, because of the synergy between HCI and Gagné’s theories relating to the conditions of learning.

More recently, in 1988, psychologist John Sweller made a link between cognitive load and earlier research relating to the capacity of working memory. Sweller’s research in

mathematical problem-solving lead to the formulation of CLT. This model has evolved over the last thirty years and has been adapted to investigate ways in which cognitive load can be measured in various domains including health-professional education (van Merriënboer and Sweller, 2010).

1.5.3. Cognitive Load Theory

CLT is based on the cognitive psychology of George Miller (1956). Miller recognised that the human short-term memory has limitations and concluded that it could only hold between five and nine pieces of information at the same time. It is now accepted that human cognition features a limited working memory, but a relatively unlimited long-term memory (van Merriënboer and Sweller, 2010). The long-term memory contains collections of items organised into schemata (the concept first theorised by Piaget in 1929).

In computer-based learning and m-learning, an acknowledgement of cognition is integral to the design of learning materials. Studies have shown that some types of media and media-delivery methods may hamper learning because of their negative effect on cognitive processing. Goal-orientated learning tasks and any learning content that splits the attention (such as presenting multiple sources of information that require mental integration by the user) may cause an information overload (Sweller, van Merriënboer, and Paas, 1998; Liu, et al., 2012). Animations are also often cited as being responsible for overloading the cognitive processing power of the learner (Boucheix and Schneider, 2009), particularly when they contain so-called *seductive detail*. This refers to information that may be interesting, but is not relevant to the learning task at hand (Park, et al., 2010). Touch-screen devices not only have the ability to show content with seductive detail but also present the user with audio, text, imagery and a high degree of tactile interaction via the touch-screen display. A touch-screen interface is unique to these devices and quite unlike the traditional mouse input method. All of these features may be capable of contributing to the cognitive load imposed on the learner and are critically evaluated in the literature review from page 60.

1.5.4. Human-Computer Interaction

Cognitivist theory is very closely linked to the study of HCI. In the words of Hurtienne (2009, p. 13) “The advent of computers influenced cognitive science and cognitive science influenced how computers were built”. HCI can be traced back to the 1960s when the advancement of computing technology resulted in what was termed “the software crisis” (Haigh, 2010). This predicament was caused by the fact that many new software applications were required for increasingly complex computer architecture. At the time, computers were not provided with a graphical user interface (GUI) (such as Microsoft Windows or Mac OSX) with which the user could interact with the device. However, in 1972 the Xerox Corporation released the Alto system, which used a new approach. This model was the first computer to feature a bit-mapped (pictorial) display, the use of graphical representations of windows to represent folders and was the first computer to use (and lead to the invention of the term) *icons* (Cruzi, 2003). This concept was quickly adopted by the Apple Computer Corporation (1983) in their first mainstream computer the *Lisa* and shortly afterwards the *Macintosh* (1984). Microsoft followed suit in 1985 with *Windows*.

With the advent of the GUI and the use of input devices such as the computer mouse, the focus of HCI shifted to the design of the interface, with particular reference to usability. The approach taken was very much based on the cognitive sciences that had developed over the 1970s following the *cognitivist revolution* of the previous decade. One of the key figures in this movement was Christopher Longuet-Higgins, co-founder and director of the Institute of Cognitive and Information Sciences in Sussex, UK. In this role, Longuet-Higgins encouraged a collaboration between various disciplines including computer science, language, neuroscience and experimental psychology leading to the new discipline of Cognitive Science (Nuyts, 1990).

The reason that cognitive theory is relevant in HCI and especially relating to mobile devices and m-learning is that a smartphone or tablet can be defined as a cognitive artefact – a human-made tool (computer) designed to support mental activity (Terras and Ramsay, 2012). HCI looks at various factors relating to the use of computers such as:

- cognitive work analysis (making socio-technical systems easier to understand)
- distributed cognition (developing technology to support human interaction)
- gesture interaction (such as the use of touch-screens, styluses and mice)
- information retrieval (browsing/searching and finding data, documents and files)
- mental modelling (how the user perceives the structure and function of a device)
- visual representation (such as symbols and icons)

These factors highlight the fact that m-learning sits at the human/computer interface where Longuet-Higgins' fields of study converge and overlap. These fields traditionally centre around computer science, but from the learning and teaching perspective, the other areas of communication theory, language and cognitive psychology come into play. In addition to software, there are also hardware concerns. Early research in the field often focussed on the instrumentation. Kjeldskov (2003) discovered that 61% of the existing research into mobile HCI looked at engineering or re-engineering components, developing new parts and considering the properties of a product. These elements do not, however, focus on the end-user. Human/computer interfaces must allow for user-interaction where failure to understand the needs and limitations of the user can lead to the failure of an entire system.

All of these HCI factors need to be considered when designing learning materials, and the process is often based on an approach known as *instructional design*. This theory, also developed by Robert Gagné (Gagné, Briggs and Wagner, 1998), has its roots in the 1940s when Gagné was involved in the design of training materials for soldiers during the second world-war. Instructional design offers a structured approach to the creation of learning materials and is based on Gagné's cognitive model relating to the mental processing of information. There are some criticisms of Gagné's original model, particularly the fact that instruction is passive rather than active, and therefore does not encourage the learner to engage in the cognitive process of making meaning (Merrill, Li and Jones, 1991). Furthermore, the mode of delivery may be inflexible in that the materials are predetermined and fixed and are not self-adapting in reaction to the input from a learner. However, these shortcomings have largely been addressed over the last thirty years on account of the fact that the student is no longer required to be a passive

observer. Digital technology affords the learner the opportunity to interact with learning-materials. Instructional design theory has been expanded (into what is known as second-generation instructional-design) to take into account the increasing use of technology, such as hypermedia and interactive software design in education. Merrill, Li and Jones (1991) recognised the need for this adaptation when technology-based delivery systems (web-based learning) began to emerge. They explain that the second generation of instructional design recognises the fact that the learner can now interact with the learning activity, and that the software can dynamically customise the content in reaction to the learner's needs. To achieve this "dialogue" between human and machine, and to ensure that learning is designed to effectively facilitate the formation of schemata, cognitive theory is now firmly embedded into the instructional design of e-learning materials and has developed into the *Cognitive Theory of Multimedia* (Mayer, 2009; 2012). This topic is covered in the literature review chapter on page 69.

1.6. METHODOLOGICAL CONSIDERATIONS

In the previous section relating to professional practice (page 6) I indicated that my experience of research was entirely quantitative and experimental in approach. However, in any new research it is important to consider the question being asked, and the nature of the data being collected when defining the ontological and methodological approach to be used. Waring (2012, p.16) identifies four building blocks of research, namely ontology, epistemology, methodology and method. These concepts are summarised by Arthur, et al. (2012) by four key questions:

“What is the form and nature of the world?” (ontology)

“How can what is assumed to exist be known?” (epistemology)

“What procedure or logic should be followed?” (methodology)

“What techniques of data collection should be used?” (methods)

The main underpinning philosophical traditions behind research are broadly divided into the constructs of positivism, post-positivism and constructivism (interpretivism) (Rugg and Petre, 2007; Arthur et al., 2012). In educational terms, these traditions are aligned to learning theory, from the experimental (positivist) approach taken by the founders of behaviourism such as Pavlov, Thorndike, Watson (1931) and Skinner (1988), through a post-positivist, cognitive approach to the constructivist perspective used throughout modern learning and teaching (Piaget, 1969; Vygotsky, 1986). There has been a paradigm-shift away from positivism in education and research, not just because of the cognitive revolution, but also following our better understanding of quantum physics, where a degree of uncertainty is now accepted as unavoidable (Heisenberg, 1998). Science now often takes a post-positivist approach, whereby quantitative measurements deal in probability rather than certainty (Salkind, 2008). Bearing these philosophical traditions in mind, the ontological and epistemological perspectives to be used in this study were largely dictated by the research question.

1.6.1. Ontology

Cognitive load is quantifiable. Quantitative, empirical studies (i.e. where measurements are made) normally require the ontological assumption that the world contains order, structure and obeys certain universal laws (such as the laws of thermodynamics or motion). These assumptions are needed to substantiate what Hedges (2012) calls the “logic of enquiry” where the research design is aligned to the research problem, but also to the data collection method. The quantitative nature of this study places it towards the positivist end of the philosophical spectrum. However, when measuring cognitive activity, there are both subjective and objective approaches that may be used in data collection (these are more fully evaluated in Chapter 3). Objective approaches include pre and post-testing of the participants, which provides numerical data. However, subjective methods such as those used in this study are often contingent on self-reporting of task load (Hart and Staveland, 1988) and rely on the internally-constructed perceptions of the participants. This Likert-scale data, although numerical, reflects the experience of the individual and therefore shifts the ontological perspective a little towards the constructivist tradition.

1.6.2. Epistemology

Epistemology can be defined as a study of the nature of knowledge (Martin, 2014). Just as ontology is consistent with philosophical tradition, epistemology also has two main branches that align to the poles of constructivism and positivism. These are identified by Martin as the epistemology of the *a priori*, which takes a rationalist stance, and the epistemology of the *a posteriori* (in Martin’s words “knowledge based on sense-experience) which was developed in the late seventeenth century by empiricists such as Thomas Hobbes and John Locke.

The *a priori* model of epistemology takes a Cartesian perspective, which centres on the formation of beliefs through rational intuition. Here there is an understanding that human perception can be flawed, knowledge is innate, and that only clarity of thought can lead to a true understanding of the world (Descartes, 2011). Although it is demonstrably true that human perception can be flawed, I have already indicated in the introduction that there are empirical arguments against the premise that all knowledge

exists *a priori* of experience. The philosopher David Hume (2011) made the proposition that some knowledge can be classed as a *matter of fact*, whereby something can only be said to be true if it is observed. Positivists, for example were reluctant to accept atomic theory at a time when atoms could not be directly observed. From an ontological perspective, post-positivists would object to this view, because it is possible to demonstrate the existence of atoms without having to physically see them (Singh, 2005). However, in Hume's definition, he continues to explain that it is not self-contradictory to suppose that there might be an alternative to the phenomenon that is being observed. Hume's idea that empirical observation leads to knowledge and that propositions may be shown to be false by further observation, is one of the underpinning tenets of scientific research and leads to the *a posteriori* branch of epistemology.

The *a posteriori* episteme holds that concepts only arise from experience and lies in the empirical tradition of Aristotle who is credited with the slogan "nothing in the intellect not previously in the senses" (Martin, 2014, p. 108). This saying refers to the belief that knowledge can only be gained through direct sensory experience. The concepts involved in empirical epistemology are fundamental principles that do not rely on rational intuition and require minimal personal interpretation. They can be measured. However, there is still reason for caution. Observational data sometimes rely on perception and interpretation by the individual; there can be errors in measurement, misunderstandings about the difference between correlation and causality and other confounding factors. This results in the necessity for a degree of scepticism. Martin identifies that scepticism can be constructive, whereby the sceptic seeks to provide an explanation or to provide an account of something that was formerly misunderstood or unclear. Hume realised that this was related to causality. He argued that the contiguity of events does not necessarily mean that one event caused another, and even if so, it might not hold true for future occasions of the same event. This realisation is now also firmly entrenched in the scientific method.

In summary, taking all of the above into consideration, and applying the theory to my study, it seems that there are two epistemological approaches that could be used. A subjective, inductive paradigm or an experimental, deductive approach. When undertaking a scoping review of the literature for this study, I noted that there had

already been a number of studies examining m-learning from an inductive, qualitative approach. Mayer (2009) identifies the importance of such learner-centred approaches when using multimedia materials, so this is an important area to evaluate. Many of the published papers that I sourced related to the student experience, seeking to gain an understanding of the feelings and opinions of learners in relation to the use of information technology (Dix, 2004; Kukulska-Hulme and Traxler, 2005; Zhang and Galetta, 2006; Sun, et al., 2008, Coulby et al., 2011; Gikas and Grant, 2013). Studies such as these often employ a hermeneutic approach (requiring interpretation) and are useful in gaining an understanding of a phenomenon. Although my initial literature search identified many qualitative studies that typically showed a relationship between student experience and e-learning, there seemed to be very few studies that looked at m-learning outcomes using an experimental approach. In particular, there were none that appeared to quantify cognitive learning outcomes when comparing m-learning with traditional paper-based materials. This apparent gap suggested a contribution could be made to knowledge in this area, and lead to my formulation of the research question stated at the beginning of this chapter. In this case, the research question defines the approach to be taken. Because there is a pre-existing question and a number of hypotheses to be tested, the epistemology into which it falls is deductive. It requires an objective, top-down strategy where there is no presupposition about what the result may be and there is intended to be a creation of *a posteriori* knowledge. In this model, the researcher strives to maintain an objective detachment from the data-collection process and seeks to produce empirical results that may be generalised to a wider field. Critics of this method might say that it lacks authenticity, the participants are placed in a location and asked to perform tasks under conditions that may differ from their natural environment. Also, the data collected may not provide a deep understanding of the phenomenon being studied (Denzin and Lincoln, 2005) My counter-arguments are that any flaws in authenticity can be largely controlled-for and although the nature of the data collection tools chosen will provide largely numerical data, there are subjectively-derived data that reflect human perception. I propose that this will allow an understanding of the learning process in the context of my study using the theory of cognitivism.

M-learning differs from traditional learning in two significant ways, firstly, the fact that a mobile electronic device is used, and secondly that the teaching materials often contain a high degree of multimedia content. In the previous section, I explained how HCI is commonly evaluated using a model based on the cognitive psychology of Gagné, but cognitive load has also been identified as a method of assessing multimedia-learning by educational psychologist Richard Mayer. Mayer (2012) makes the case that these materials must be proven to aid learning. Research can be of benefit here, in looking at how the style and delivery of learning materials can affect cognitive load in the learner. CLT was first developed by John Sweller (1988) and has been well investigated in the field of psychology and multimedia learning (Sweller, 1989; Sweller, 1994; Brünken, et al., 2002, Brünken, Plass and Leutner, 2003; Chandler, 2004; Paas and Sweller, 2011; Vogel-Walcutt, et al., 2011; Sweller Ayres and Kalyuga, 2011; Paas and Ayres, 2014) but to date has not been fully employed in the context of mobile touch-screen devices.

1.7. MOBILE LEARNING AS A RESEARCH TOPIC

My pedagogical interest not only lies in the learning of medically-related subjects such as human anatomy, but also how technology has the capability to bring inaccessible subjects to a wider audience. Inaccessibility may be interpreted in various ways. To use the example of human anatomy, a student wishing to learn about this topic may not have access to materials such as anatomical specimens or models. Inaccessibility may also refer to understanding; some subjects may be described as “impenetrable” due to their inherent complexity (Mayer, 2012). Inaccessibility may also apply to human contact with peers and educators (Martin and Ertzberger, 2013). This section identifies the scope of m-learning. The benefits and limitations of m-learning are evaluated to justify the research question, and the underpinning educational theory that informs m-learning (particularly relating to the cognitive domain) is identified (Bloom, 1956; Dave, 1967; Anderson and Krathwohl, 2000; Krathwohl, 2002).

M-learning has been defined in many ways, but in essence, it involves a widening of the traditional learning-space by the use of wirelessly connected mobile technology such as smartphones (Pachler, Bachmair and Cook, 2010). Recent advances in technology have

equipped mobile devices with touch-screen capability, an increasing range of sensors and high-quality cameras that can be used for instant live communication (Khan, Yang and Arshad, 2013). These features combined with portability and fast network connectivity open up new possibilities for how mobile devices can be employed as learning tools in areas such as fitness, health, student collaboration and activities undertaken in informal learning spaces (Young, 2010). New branches of e-learning such as m-learning and ubiquitous “anywhere/anytime” learning (u-Learning) (Shin, et al., 2011) have been defined to recognise the educational potential of this evolving technology. M-learning is a subset of computer-based learning, which in turn, is a subset of the larger field of distance-learning. However, like other computers, mobile devices can also be used on campus and due to their ubiquity, can also be used in the classroom setting under certain circumstances (Manuguerra and Petocz, 2011; Clarke and Luckin, 2012). Figure 1-3 is a representation of the boundaries of m-learning as a subset of CBL. The section areas are not intended to represent proportional scale. In this diagram, it is recognised that online learning can be undertaken on mobile devices as well as personal computers and that both of these modalities can be used both in the classroom and at a distance.

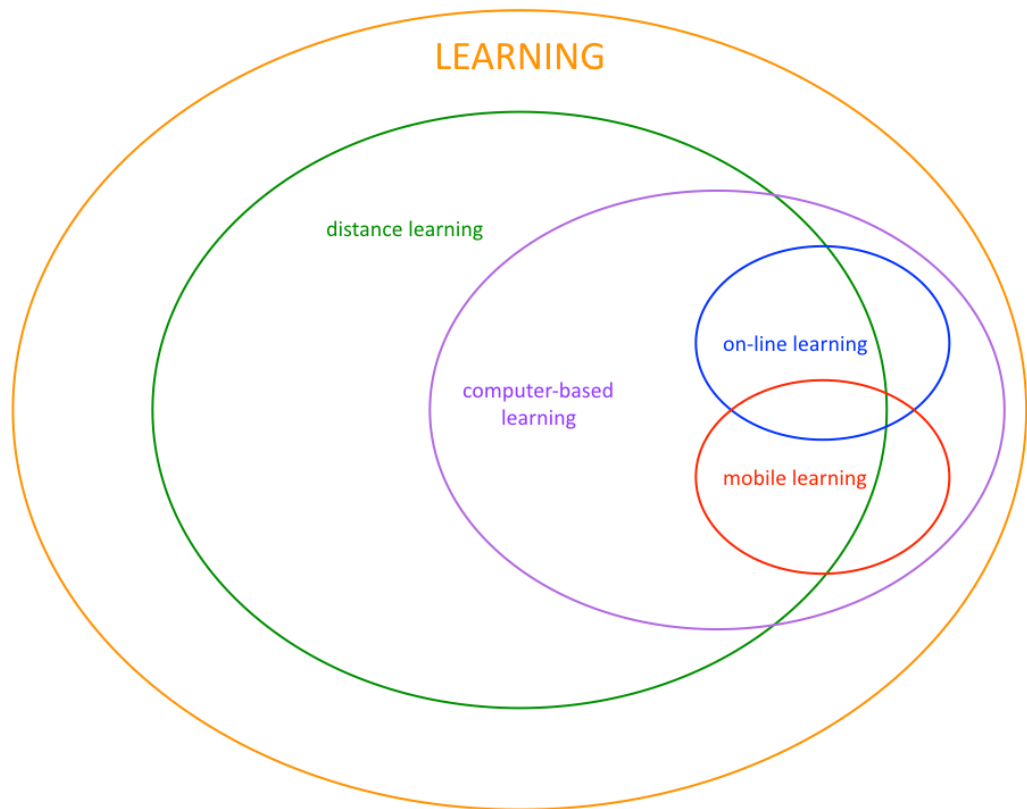


Figure 1-3: m-learning as a subset of CBL, note that the section-sizes are not intended to be representative of the scale of employment.

E-learning is a term that defines the educational use of digital (electronic) technology. The term was introduced in 1998 by Cross (2004), and there are many definitions in the literature.

The European e-learning action-plan of 2001 provides an early definition:

“the use of new multimedia technologies and the Internet to improve the quality of learning by facilitating access to resources and services as well as remote exchanges and collaboration”. (Commission of the European Communities, 2001, p.2)

More recently, Garrison (2011, p.2) has defined e-learning as:

“electronically mediated asynchronous and synchronous communication for the purpose of constructing and confirming knowledge”.

Many courses now augment their face-to-face lectures with electronic materials in a blended or resource-based learning approach (Gikas and Grant, 2013). Universities have embraced this concept; the 2015 NMC Horizon Report observed that many institutions had upgraded their wireless bandwidth to create “smart-rooms” to permit remote collaborative communication, and libraries have included spaces for the use of e-books and online resources in their design. However, one of the major advantages of e-learning is its application across a geographically-wide community of students (Traxler, 2009; Zhao, 2011). This location-independent feature of e-learning is ideally suited to the use of devices such as smartphones as they are connected to mobile networks in addition to wireless networks (Vavoula, Pachler and Kukulska-Hulme, 2009). Definitions consider m-learning from either a technological device-based perspective or a learner-centred perspective that focusses on the spontaneity, informality and intimacy presented by social collaboration (Sølvberg and Rismark, 2012). The definitions range from the very brief

“digitally-facilitated, site-specific learning” (Laurillard, 2007, p.156)

to more complex interpretations such as:

“the exploitation of ubiquitous handheld technologies, together with wireless and mobile phone networks, to facilitate, support, enhance and extend the reach of teaching and learning” (Hashemi, et al., 2011, p.2478).

This second definition sums up most of the key concepts of m-learning, particularly the recognition that m-learning typically involves interaction with hand-held devices, and that these devices are now ubiquitous. Other authors identify laptop computers as m-learning devices (Sung, Chang and Liu, 2016; Sarab, Elbasir and Alnaeli, 2015), but this does not appropriately reflect the *anywhere, anytime* principle of mobile technology (Perry, et al., 2001; Traxler, 2007). Laptop computers tend not have the universality, portability or cellular network connectivity required for truly location-independent

learning. M-learning is therefore characterised by the use of mobile devices such as smartphones and tablet PCs, particularly tablets having a pocketable form-factor such as the *iPad Mini* (Apple Inc.) or the *Google Nexus 7*” device (Google Inc.). These devices tend to be readily to-hand and allow spontaneous and continuous access by the learner. Network subscriptions for computing devices provide a useful snapshot as to their penetration into the general population. Laptop and personal computers are fewer than 250 million globally, whereas mobile device subscriptions exceed 2.6 billion (Ericsson, 2015).

Mobile devices were recently identified in the NMC (New Media Consortium) Horizon report (2016) as a technology that has been widely adopted in teaching and learning. Many educational establishments have incorporated technology into their learning spaces to improve mobility and flexibility in learning. The NMC Horizon Report (2016) observes that mobile devices have been introduced to take advantage of technological advances in areas such as data analytics, games, simulation, social media and communication. In addition, the user-familiarity exhibited by the current generation of school leavers, who have grown up with the technology – appears to foster student engagement (Manuguerra and Petocz, 2011; Clark and Luckin, 2012; Martin and Ertzberger, 2013; Johnson et al., 2015). In the private sector, companies such as Apple, Google and Samsung have moved into the educational market with mobile distance learning (DL) environments such *iTunes U*, *Google for Education* and *Samsung Smart-School*. These resources, particularly from Apple and Google have expanded existing commercial-content delivery systems into educational solutions, often provided free to the end-user. Apple’s iTunes store has distributed digital media since 2003 and branched into the delivery of University lectures in 2007. iTunes U (University) now offers over half a million different lectures from over 1000 institutions including the world’s most prestigious universities (Apple, 2013; Watson, 2013).

1.7.1. Touch-screen Devices in the Context of e-learning

In the early days of the World Wide Web it was noted that the dynamic flexibility and nonlinearity of hypermedia (text and images with interactive navigational links) allowed the creators of learning materials to present information in what McGuire (1996) calls “a

conceptual web that mirrors some of the associational power of human memory”. This link between human memory and computing is an association that dates back to the cognitive revolution and the works of Robert Gagné which were evaluated in the previous section beginning on page 16.

McGuire examined the qualities of web-based hypermedia that can be exploited to foster learning. There was a fundamental distinction between hypermedia and traditional text in that the hyperlinks embedded within the materials allowed the learner to take their own path. In doing so the learner was able to develop a personalised knowledge-base, using a constructivist approach to learning as pioneered by Piaget and Vygotsky (Holmes and Gardner, 2006). McGuire’s paper identifies that the rich diversity of materials available online allows the learner to compare their personal viewpoint to those of other authors and construct their own knowledge in a way that fosters higher-order thinking. These diverse materials are collectively described as *multimedia*, a term that was first used in the art-world to describe an installation having additional media, typically sound, incorporated into the work. In the context of learning, Psychologist Richard Mayer (2009) explains that the term *multimedia instruction* can be used to describe any presentation involving the use of text and graphical content intended to foster learning. Multimedia content can range in complexity. Simple multimedia includes content such as found in e-books, where traditional text and image-based content may be navigated in a non-linear way (Younus, et al., 2011). More sophisticated multimedia is characterised by the functionality of mobile applications (apps) that can contain audio, video and augmented-reality elements - where digital information is overlaid onto a live image created by the camera on the device (Kipper and Rampolla, 2012). The use of dynamic multimedia is, therefore, one of the principal differences between computer-based learning and traditional paper-based learning.

The Open University (OU) recognised the value of multimedia in distance-learning from the outset, particularly in the accessibility offered by television broadcasts. The OU lecture-style broadcasts ended in 2006, but recognising that accessibility remains one of the fundamental requirements of distance learning, the OU are now looking towards mobile devices as content delivery systems. Nicholas Watson, project manager for m-learning, has been involved with the digitisation of their courses for mobile device use.

Watson summed up the unique benefit of accessibility in m-learning in an iTunes U webinar for Apple developers in 2013. He commented that;

“you have to get yourself into the mind of the remote student. You must remember that your student may not be at a desk or in front of the TV, they could be on a bus or train or a park bench in the English drizzle rather than chained to a computer” (Watson, 2013).

Accessibility in e-learning leads to many advantages. From a Vygotskian perspective, it allows collaboration and the construction of learning between individuals who may be geographically remote. Kearney et al. (2012) emphasise that m-learners can use the socially-interactive environment offered by mobile devices (and their associated applications) to construct learning by the sharing of digital content across “time and space”. By this, the authors are referring to the authenticity that can be realised when learning is situated (for example in the field) and the collaboration that can be fostered by data sharing and communication. These advantages hinge on the accessibility and availability of other users in real-time, something that is uniquely afforded by devices on a mobile network as they can access the network from remote locations. Laptops are unable to offer this location-independence unless fitted with a *subscriber identity module* (SIM) or have access to a personal Wi-Fi modem. Without such features, computers are restricted to a wireless network connection via local routers that may not be available in many locations.

The fact that mobile devices augment human cognition was identified by Terras and Ramsay (2012), who noted that (unlike traditional mobile phones) smartphones are not solely for communication. The devices supplement cognition at several levels. In the most basic sense they provide calculation functions, memos and reminders. At a deeper level, they provide collaboration via mobile internet access and distributed cognition whereby knowledge is offloaded onto artefactual objects. (Pachler, Bachmair and Cook, 2010). Terras and Ramsay explain that if these devices are to augment learning, it is necessary to consider the underpinning psychology relating to the cognitive capability of the learner. Psychology offers what the authors describe as a “methodological toolkit”, looking at new

behaviours and challenges that have been thrown up by the new context of m-learning. These are discussed in the following section.

1.7.2. The Benefits of m-learning

Extending the Curriculum

The immediate fore-runner of today's touch-screen phones and tablets was the Microsoft Tablet PC. The value of the Tablet PC as a learning device was identified by Twining and Evans (2005) who undertook a very thorough comparison of Tablet PCs and other formats of computer used in the classroom. At the time of the study, tablet devices were very expensive and the research performed was sought to justify the additional cost. Data was collected from a literature review and 12 case studies, each undertaken in a different school. There were areas where tablets offered benefits over the traditional laptop or PC. A primary finding was that the students had a high degree of empathy with the tablet devices. This positive user-response was thought to be related to the portable size, form, functionality and intuitive interface. From a pedagogical viewpoint, it was found that the touchscreen devices also provided some unique advantages including the provision of extended access to the curriculum for students with special needs, providing a new medium for artwork, supporting learning outside the classroom on field trips and for homework and encouraging collaboration and independent research.

Pachler, Bachmair and Cook (2010), further investigated this theme by examining the role of mobile devices in situated learning. Situated learning takes a Montessorian perspective, in which learning is considered to be located within an authentic activity, context and culture, and relies on meaning-making. The authors acknowledge that the school classroom may be unsuited for this type of learning as it is constrained regarding the range of situations it can offer. In a similar way to the TV broadcasts formerly used by the Open University, mobile devices can help to bring socio-cultural, geographic and communication-based contexts into the classroom, whereby media becomes a cultural resource inter-related with child development. More recently-published work has recognised that smartphones and tablets can aid productivity in teaching and bring computer-based learning out of the IT lab and into a more casual environment. Shortly

after its release, Young (2010) described the first iPad as “bean-bag” friendly. He likened its unobtrusive form-factor to a sketchpad rather than a computer – encouraging informal learning in spaces other than the classroom.

In their 2012 report, the Technology Enhanced Learning (TEL) Research Programme highlighted an interesting finding relating to mobile devices. Allowing students or schoolchildren to bring smartphones into the classroom has traditionally been thought to be disruptive (in a negative sense). But there is increasing evidence that these devices have benefits when used in a way that facilitates learning. In the report identified that these benefits include:

- extending classroom activities,
- allowing the students to continue working at home,
- providing alternative perspectives from sources other than their teacher/lecturer,
- allowing children to take notes or collect data away from the classroom, and
- encouraging the use of community learning and social interactions that would not be available in a school or lecture hall.

These findings are in strong agreement with Kearney et al. (2012), who outline a pedagogical framework for m-learning from a socio-cultural perspective. The authors suggest that m-learning has the potential to add a high degree of authenticity to the learning environment, from both a task level and process level. *Task authenticity* relating to the realism of learning tasks (i.e. offering problems that are found in the real world) and *process authenticity* referring to the realism of learner practice. The rich context of such tasks can be supported by the features of their mobile devices – such as data capture and geolocation via global positioning satellite (GPS).

Another unique advantage of the touch-screen tablet has been identified by Manches (TEL Report, 2012), who points out that young children can find it difficult to manipulate the traditional user input devices for digital technology, namely the mouse and keyboard. Touch-screen devices make interaction easier for this group and the gesture based navigation - such as swiping or pinching - offers new ways to manipulate digital content. Perhaps this is a contributor to the fact that one-in-ten pre-school-age children and one in

three children aged 5-15 now possess their own tablet (OFCOM, 2015). From an educational perspective, the physical interaction that is uniquely afforded by a touch-screen may have more benefit than ease-of-use or user-engagement. Manches (2012) states that there is increasing support for the notion that physical movement, such as gestures, are strongly linked to our cognitive processes. For example, children are able to use gestures to indicate what they are thinking before they have learnt to talk. In educational psychology, this development stage is what Jerome Bruner (1960) defines as the *enactive* stage of learning and Piaget (1969) classifies as the *sensorimotor* stage. It therefore seems reasonable to hypothesise that touch-screens could enhance learning due to their capability to capture and interpret such gestures.

User Engagement and Empathy

Distance learning environments are technologically equipped in a way that allows them to offer more than a simple server of text documents, and it is crucial that educators can use the devices in ways that foster engagement. An example of the value of user-engagement is demonstrated in another early-adoption study at Macquarie University (Sydney, Australia). The initiative was undertaken to enhance modules teaching probability and statistics. Lecturers Manuguerra and Petocz (2011) noted that traditional slide presentations (PowerPoint (Microsoft Inc.)) had resulted in disengagement by students who may not have had a direct interest in the topic. Part of the issue was considered to be due to the static nature of the slides, but also due to a lack of experience in taking notes during the lecture. With the introduction of iPads into the classroom, it was possible to shift the content onto the devices, using more “lively and spontaneous” presentations and allowing real-time annotation of the content. Manuguerra found that this strategy resulted in manifestly higher levels of student interest and participation, and gave the students reassurance that all of the content and the annotated notes recorded during the session would be available on their devices for revision purposes.

In the UK, the 2012 report from the Technology Enhanced Learning group (TEL) contained a similarly convincing case-study that showed an empathy between students and mobile devices and provides a persuasive example of where mobile technology has been used to great effect as a tool to foster student engagement. The research in question was

undertaken by the Essa Academy, situated in a multi-ethnic suburb of Bolton. In 2008, the school had a poor academic reputation with only around 25% of the pupils achieving five GCSE passes. In an attempt to raise standards all 900 children were provided with an *iPod Touch* (Apple Inc.). Initially dismissed by some detractors as a gimmick, the devices were instrumental in improving standards. They were used to facilitate communication between pupils and staff, monitor the progress of the learners and to encourage online research and the use of educational software (iTunes U). Over two years there was a marked improvement in examination results with 99.5% of pupils attaining five A* to C-grade passes in their GCSEs. The school's principal, Jeff Ellis, stated that the innovative use of technology had been a driver in this improvement and that the devices had proved to be motivational, empowering and had removed the limits to learning (Clark and Luckin, 2012). Similar improvements in learning outcomes have also been demonstrated when using iPads in the classroom (Subramanian, 2012).

The cases discussed so far have shown examples where the unique features of m-learning have successfully augmented the educational process. However, there are also some limitations to m-learning which, if not addressed, may be counterproductive to effective retention of knowledge. These are evaluated in the next section.

1.7.3. The Limitations of m-learning

Equipment Limitations

Models of quality in m-learning typically feature technical factors such as functionality, hardware performance, interface-usability and device connectivity (Sarrab, Elbasir and Alnaeli, 2016). Such studies tend to examine the causal relationships between technological features and learner satisfaction. There are studies to show that mobile devices may be lacking in some areas, compared to other types of computing device. In a paper of 2014, Souleles et al. took a qualitative phenomenographic approach in exploring users' perceptions of m-learning with *iPads*. The study found that in some learning-situations laptop computers have been shown to offer user-perceived advantages over mobile touch-screen devices. The participants, forty design-students, recognised certain benefits in using the tablets, but there were some reservations. The lack of a physical keyboard was a concern, with some participants stating that they would much prefer a laptop for any writing activity. Limitations in software and processing power were also identified, one participant stating that any sketches made on the touch-screen would then need be transferred to a personal computer for post-processing in *Photoshop* software (Adobe Inc.).

Due to the physical size of mobile devices, one of the principal factors that can affect the learning experience is screen-size. This factor is less applicable to tablet devices, as the screen size may be larger than that found on some laptop computers (for example an iPad Pro (Apple Inc.) has a screen size of 12.9" compared to a MacBook laptop (Apple Inc.) having a 12" display). Typical smartphone screen sizes range between 3.5" (iPhone 4) up to 6" (Motorola Moto 6) all of which are physically much smaller than a normal PC monitor (typically around 24"). There have been a number of studies looking at the effect of screen size on the user experience, particularly in the field of web-design. Here it is desirable to use responsive page layouts (interfaces) that adapt their configuration according to the device and screen size in use (Bohyun, 2013; Mohorovicic, 2013; Snell, 2013). Following the introduction of smartphones, Findlater and McGrenere (2008) performed an empirical study to compare adaptive interfaces for small screens. The study lacked ecological validity in that the participants were only required to interact with a PC

monitor rather than a smartphone. However, the experiment attempted to simulate two screen-sizes by presenting content at 800x600 pixels (typical screen resolution of a PC at the time) and in a smaller window of 240x320 pixels (simulating a mobile device). The task, which involved making menu selections, was automatically monitored by the software and the results showed that participants were significantly slower when using the smaller screen configuration.

The same principles apply to the range of screen-sizes on mobile devices. Raptis et al. (2013) conducted a quantitative study in which participants were required to interact with smartphones having screen sizes of 3.5", 4.3" and 5.3". The experiment was controlled for brand, attractiveness and application by using Samsung devices that were equipped with the same operating system. The findings suggested that devices having a screen size of greater than 4.3 inches improved the efficiency of the device when used for activities such as web browsing. Screen-size factors were assessed using pairwise comparisons of participant completion times for certain activities and seemed to be related to tasks that were not easy to complete on the device, for example where content scrolling was required. There were certain limitations to the study, which did not gather any data at screen sizes associated with tablet devices, and failed to collect information on a representative cross section of everyday smartphone tasks (such as map navigation) but the results suggested that task-efficiency increased with screen-size.

User Resistance

One of the first universities to trial the iPad as a tool to deliver educational content was Stanford University (California, USA). Stanford, the birthplace of Google and Yahoo, might be expected to be a technologically-astute campus, but when iPads were introduced (as an experiment intended to reduce the excessive use of printed materials) many students were reluctant to accept them. In some classes, 50% of the students stopped using the devices altogether (Keller, 2011). However, the reasons for poor adoption by the students may not have been a fault of device or software design. There were network speed issues (students on average carrying more than two internet connected devices on campus placed a strain on network bandwidth), and it was identified that staff were slow in exploring the educational potential of m-learning. Such educator-resistance to learning

technology is a recognised phenomenon. Kukulska-Hulme (2012) identified a survey undertaken in the USA revealing that the range of devices and technologies used by students in their personal lives are seldom used by the educators responsible for their teaching provision. The reason, in this case, lay in the fact that those responsible for curriculum-design and teaching-delivery are not necessarily conversant with the new technology, and therefore may not be able to visualise how this technology can be used in a pedagogical context. Kopcha (2012) also highlighted issues whereby teachers perceived the integration of learning technology to be a burden on their time and identified a barrier caused by lack of training in troubleshooting equipment faults. These findings agree with a similar, earlier study by Lim and Khine (2006) who additionally found that there were both intrinsic and extrinsic factors that may form a barrier to the integration of technology. Extrinsic factors included lack of access to equipment and lack of support; intrinsic factors took into account an underlying lack of belief that technology could enhance learning. Even when educators are technologically adept and willing to engage with m-learning, it does not automatically guarantee successful outcomes.

Laurillard (2008a) points out that it is important that education is not led by technology and, as educators, we need to be thinking about how the technology can best work for us. Laurillard uses podcasting as an indicative example. Technology has given us the ability to compress audio into easily downloadable files, allowing students to listen to lectures at a time and place of their own choosing, but Laurillard makes the point that

“...no one ever suggested that the reason why education is failing, is that learners do not have enough access to people talking to them” (Laurillard 2008b p.139).

This view is echoed by Chandler (2004), who states that instructors frequently make the crucial mistake of allowing technology to dictate the way that they create the learning experience rather than the other way around. Materials must be created with the user in mind rather than to showcase the capabilities of the hardware.

Cost

Human resources are expensive and traditional education may not be as cost-effective or

efficient as m-learning (Arrigo et al., 2013). Early examples of e-learning included the adoption of CD-ROM (Compact Disc Read-Only Memory) as an inexpensive medium to replace physical volumes of text. More modern examples include educational apps for tablets and freeware education platforms such as *Moodle*. From a hardware perspective, m-learning also has cost advantages. Tablet computers tend to be considerably more affordable than most home computers or laptops (Hashemi et al., 2011). However, Littlejohn (2003) recognised at an early stage that e-learning materials can be expensive to produce. This is particularly the case if the materials are designed to be accessed as mobile applications (apps) as these have long production times and may require the services of third-party developers. Software programming can be particularly time-consuming, and although the exact timing can be difficult to ascertain, one survey suggests that to create one hour of e-learning activity can require a development time of up to 220 hours, and may exceed 750 hours for complex hardware emulation (Chapman, 2006). This is largely in agreement with my own experience as an app developer. When compared to the typical development-time ratio of approximately 34:1 hours (average) for instructor-led training, e-learning could be considered to be an impracticably expensive option. Conversely, the expense may often be justified by the fact that e-learning materials are re-usable. The granularisation of e-learning materials into shareable, re-useable learning objects was first identified by Downes (2001), who, at the time, made the negative observation that global-sharing was a technically difficult objective. In modern telecommunications, improvements in network speed, data compression and server technology make this concept realisable. The ability to distribute learning objects quickly and easily over global networks to multiple users may provide the economy of scale required to justify production costs. However, this model assumes that the materials themselves are congruent with learning. If the learning-task elements are not presented in a way that is easily mentally integrated by the learner, perhaps due to hardware limitations or lack of embedded learning theory in the software design, learning will not occur (Sweller, 1994; van Merriënboer and Ayres, 2004). Learning must be strategically reconceptualised for the mobile platform, and must take into account recognised educational theory in its design and implementation. There is, therefore, the need to ensure that digital delivery systems offer a significant advantage (or at least do

not present a disadvantage) over conventional paper-based or classroom-based learning models.

1.8. A STATEMENT OF THE RESEARCH PROBLEM

The research problem centres around the fact that mobile applications are typically time-consuming to produce, and typically have a high associated production cost. Mehra (2014) explains that a mainstream mobile app requires a production team of perhaps six to ten people and can take six months or more in development time. When added to the cost of maintenance, bug-fixing updates and version-testing the cost can run into tens, if not hundreds of thousands of pounds. With educational software applications, the question is asked as to whether this investment can be justified. Does the app offer any measurable advantage in learning outcomes when compared to a traditional textbook?

M-learning is ubiquitous in offering any-time, any-place learning (Young, 2010; Sølvsberg and Rismark, 2012). Users show empathy with mobile devices (Twining and Evans, 2005, Clark and Luckin, 2012) and the portable nature of smartphones permits the extension of the curriculum into authentic environments (Kearney, et al., 2012). However, there are also limitations to m-learning identified in the topic literature. Some issues relate to hardware performance and connectivity (Findlater and McGrenere, 2008; Raptis, et. al. 2013; Saarab, Elbasir and Alnaeli, 2016) user resistance (Keller, 2011) and educator resistance (Lim and Khine 2006; Kopcha, 2012). The cost of devices can be low (Hashemi et al., 2011), but learning materials and software can incur high development costs (Littlejohn 2003; Chapman, 2006; Mehra, 2014). As computers, mobile devices also have the potential to increase the cognitive load of the user (Hollender, et al., 2010). HCI theory states that one of the primary goals of device usability is to reduce the cognitive load created by the interaction with the device (Card, Newell and Moran, 1983).

When learning human anatomy the traditional resources used by students include textbooks such as Ross and Wilson (Waugh and Grant, 2014), Netter (2014) and Gray (2009). This market is now being adapted for mobile devices with a wide range of anatomy applications being developed for use on smartphones and tablets (Lupton 2014). Authors such as Mayfield (2012) and Lewis, et al. (2014) have identified that iPads increase learner engagement and enhanced the efficiency of dissection education. There

have also been qualitative studies looking at the socio-cultural aspects of mobile apps (Lupton, 2014) and a recent qualitative analysis of mobile applications by Zydney and Warner (2015) has indicated that research into learning apps is required to assess cognitive outcomes. However, there seems to be a gap in the current literature relating to the impact of mobile apps on learning from a quantitative, HCI perspective, particularly whether the device or the interactive interface design introduces cognitive load compared to textbook learning. Factors such as screen size and screen resolution also seem to be under-researched, yet this has been identified as a possible shortcoming in studies such as Lewis, et al. (2014) who questioned the level of detail afforded by apps when learning anatomy.

In summary, there appears to be a need to assess this new mobile content-delivery method from a cognitive perspective. Learning can be assessed in many different ways, but for this study it was proposed to use performance outcome measures (summative assessment) and measurement of cognitive load. These methods were employed because they provide quantitative data, and also because the latter relates strongly to HCI (Hollender, et al., 2010). The primary aim was to determine whether the level of cognitive load generated by device interaction presents a barrier to learning, or enhances learning. If the former were found to be the case, the justification for using mobile devices over non-interactive learning materials such as books might be difficult to justify financially. Further literature-based justifications for research into this area can be found in the literature review on page 96.

1.9. RESEARCH QUESTIONS – THE PURPOSE OF THIS STUDY

The primary research question is:

Is there a statistically significant difference in the level of task load experienced by a learner when undertaking a multimedia interactive learning activity delivered by a mobile touch-screen computing device compared to that experienced by a learner undertaking an equivalent non-interactive learning activity designed to teach the same factual information?

The Secondary research questions are:

Is there a significant correlation between the task load experienced by the learner and the difference in pre and post-test scores relating to the learning activity undertaken (i.e. was there a barrier to learning)?

Is there a significant correlation between the nature of the learning activity (interactive vs. non-interactive) and the difference in pre and post-test scores between groups?

Is there a significant correlation between the age or gender of the participant and the degree of task load experienced when undertaking m-learning?

Is there a significant correlation between the screen-size or spatial resolution of the mobile device display and the degree of task load experienced when undertaking m-learning?

1.10. STATEMENT OF RESEARCH HYPOTHESES

A full list of statistical hypotheses is presented on page 122. The research hypotheses are as follows:

1.10.1. Primary Hypotheses

The primary hypotheses are:

There will be no measurable difference in net task load reported between non-interactive learners (the control group) and mobile-learners (the experimental group).

There will be no measurable difference in net pre/post test-score results between the control group and the experiment group.

In addition to the primary hypotheses, the multi-dimensional nature of the NASA TLX affords a more detailed evaluation of the learning task. This allows testing between groups for differences relating to the sub-scales of physical demand, mental demand, temporal demand, performance, effort and frustration.

Furthermore, the physical attributes of the mobile devices were also tested for correlations with the sub-scales of the NASA TLX these attributes were; device screen-size, pixel density of device screen (spatial resolution) and device type (smartphones vs tablets). All hypotheses are fully stated in section 3.6.

1.10.2. Assumptions

In the tradition of detached post-positivism, there were no formal assumptions made about the possible outcomes of this experiment. Two-tailed tests were used because statistically, the data were not constrained in a particular direction. (The direction of any statistically significant results is substantiated using descriptive statistics in Chapter 5 beginning on page 238.) However, the following assumptions could be considered:

Assumptions Relating to Cognitive Load

M-learning involves the use of a computing device; HCI is known to place a degree of cognitive load on the learner (Hollender, et al., 2010) Non-interactive learning (studying a diagram) does not involve the use of a computer. It is, therefore, a reasonable assumption that the non-interactive learner would experience a lower cognitive load in comparison to the mobile-learner. If there are more mental resources available for learning, it could be assumed that the non-interacting learners would achieve a higher mean score than mobile-learners when tested on the learning activity.

Assumptions Relating to Physical Demand

Studying a textbook diagram does not require any physical activity (other than eye-movement). It could, therefore, be assumed that there would be less physical demand placed on the non-interactive (control) group compared to the mobile-learners.

Assumptions Relating to The Characteristics of the Device Used

Hardware companies now create device-displays with high spatial resolution. The aim of these displays is to present an image where the individual pixels cannot be discerned by the human retina (Spencer, et al., 2013). This offers two main advantages to the user.

Firstly, the higher spatial resolution allows learners to discern a level of detail in photographs that is equivalent to that found in traditional textbooks.

Secondly that the amount of multimedia content that can be displayed in a given screen size will be greater, as there are more pixels-per-inch (PPI). This reduces the amount of scrolling and zooming of the image (Jones, Buchanan and Thimbleby, 2003). The assumption is, therefore, that there would be no significant difference in learning outcomes or cognitive load between learners using a mobile device or a static photograph such as found in a textbook. The range of devices used by the mobile-learners, featured various screen sizes. A large display is capable of showing more content than a smaller one. It could therefore be assumed that the users of larger screen-sizes would find the learning task easier and report less time pressure than the users having smaller screen sizes as the amount of scrolling would be reduced (Findlater and McGrenere, 2008).

1.11. SUMMARY OF CHAPTER 01

In Chapter 1, I have stated the primary aim of my research in seeking to justify the use of mobile touch-screen devices as learning tools from a cognitivist perspective.

In section 1.4, I attempted to position myself in the research in the context of my professional practice. My pedagogical field of interest is m-learning, and my professional background is in medical imaging (human anatomy). In addition to distance learning, my academic activities include writing textbooks and creating mobile applications for learning, both of which defined my choice of research topic. I introduced some concepts relating to the use of digital media in learning, and specifically how the new wave of high-capacity touch-screen mobile computing devices might offer a rich new vein of research opportunity. My previous experience of research in life-sciences has typically employed quantitative research methods that rely on statistical significance. Together with the research question, this informed my consideration of an appropriate methodology for this study.

In section 1.5, I presented a conceptual framework for the study. This uses the parallel models of *Cognitive Load Theory* (Sweller, 1989) and *Human Computer Interaction Theory* (Hurtienne, 2009), both of which can be applied to m-learning and are used in the field of educational psychology. HCI theory is particularly relevant in an educational setting as it is based on the philosophy of Robert Gagné (1985). CLT is also relevant as it builds on the work of Jean Piaget (2001) and the concepts of schemata.

I also looked at broad methodological considerations in this section and evaluated the ontological and epistemological foundations for the study. I posited that the philosophical tradition for experimental research has evolved from being strictly positivist (Watson, 1931; Skinner, 1988), and now tends to favour a post-positivist approach because science recognises probability and the fact that there is a degree of uncertainty in any measurement. I made a case against the rationalist, *a priori* episteme and justified the use of an empirical approach in collecting quantitative data. Cognitivism can be investigated using an empirical approach, as it features manifest variables that can be directly measured. It lies in the post-positivist philosophical tradition, acknowledging that

cognitivism also features internal constructions (such as comprehension or understanding) that cannot always be directly measured.

In section 1.7, I defined m-learning and evaluated the field as a suitable research topic, with particular reference to touch-screen displays. I provided a critical analysis of the historical background behind m-learning. I also explained how m-learning differs from other types of computer-based learning particularly relating to its ubiquity, access to mobile networks and the pocketable form-factor of the devices used. I critically evaluated mobile devices as learning tools, and in doing so, highlighted the benefits offered by these devices but also some limitations of m-learning that are evident in the topic literature.

This evaluation was used to formulate a research problem which was articulated in section 1.8.

Leading on from this I presented the research questions in section 1.9 and stated the general research hypotheses in section 1.10. Assumptions were stated on page 48, but I explained that the tests to be employed were two-tailed as the results could not be anticipated to be directional in nature.

2. CHAPTER 02: LITERATURE REVIEW

2.1. INTRODUCTION

The previous chapter gave an overview of m-learning and introduced some of the benefits and limitations of the genre. There is a body of evidence to support the proposition that mobile devices offer unique advantages as educational tools. They are engaging, ubiquitous, typically cheaper to purchase than home computers and are able to extend the curriculum socially, geographically, across a wide age-spectrum and to users with special needs (Kagohara, et al., 2012). The hypermedia environment provided by such devices may be used to foster a cognitive-constructivist approach and the devices tend to engage the user in a way that traditional materials do not. McGuire (1996) and Tabuenca, et al. (2015, p.54) encapsulate these points in reminding us that lifelong learning is now considered the norm and that the lifelong learner must “constantly change their learning context, location, goals, environments, and also learning technologies”. M-learning is a tool that can overcome many of the barriers thrown up by these changing contexts.

However, Human Computer Interaction (HCI) brings its own barriers to both learning and user-engagement if the equipment and software design are not aligned with cognitive theory. Cognitive Load Theory (CLT) has particular implications for medical education identified by Young et al. (2014) as being related to the high complexity of the skills to be learnt. This complexity introduces what is known as a *high element interactivity*, which can hamper cognition and provide a barrier to learning. This concept is explained in section 2.2.1.

In deciding how to approach this study, to ensure originality and to determine whether there are any gaps in current knowledge concerning the use of touch-screen devices in m-learning, a recognised model for the assessment of mobile devices was used. In 2006, Koole and Ally produced a model that they called a Framework for the Rational Analysis of Mobile Education (FRAME).

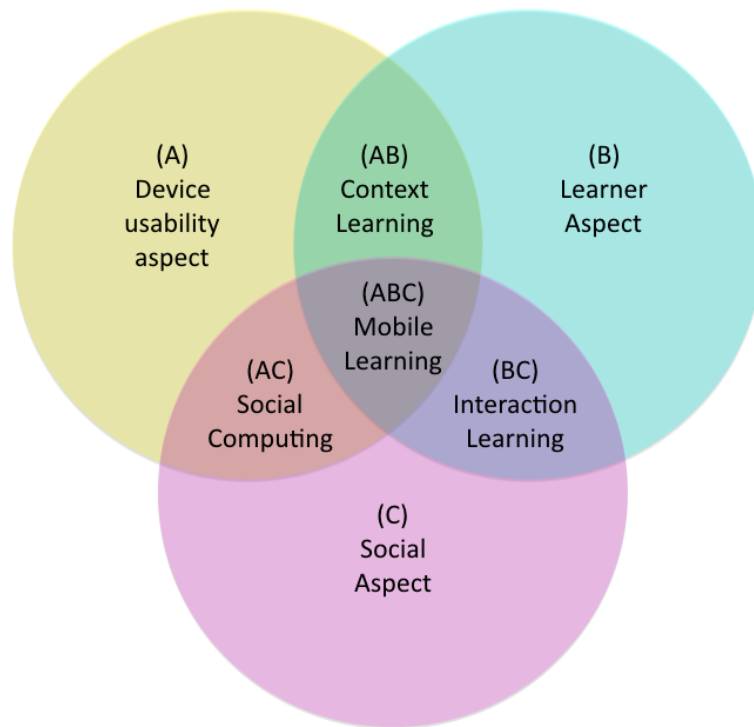


Figure 2-1: The FRAME model (Koole and Ally, 2006). The three circles of the diagram represent the device usability (A), the learner aspect (B) and the social aspect (C) of m-learning.

This model is very useful in that it provides an overview of the aspects of m-learning that may offer opportunities for research. The authors state that:

“The FRAME model is the first comprehensive theoretical model to describe m-learning as a process resulting from the convergence of mobile technologies, human learning capacities, and social interaction. It addresses contemporary pedagogical issues of information overload, knowledge navigation, and collaborative learning. It is hoped that this model will help to guide the development of future mobile devices, the development of learning materials destined for m-learning, and the specification of teaching and learning strategies for mobile education” (Koole and Ally 2006, p.1).

The overlapping intersections of the diagram highlight what the authors describe as

“synergies” between the user and the device. In other words, what the device affords the learner regarding functionality and network connectivity. The *social technology* intersection does not relate specifically to the learner, as it describes device communication technology. This area would be of interest to a mobile equipment designer, but device design is not a variable that can be manipulated by an educator. The *interaction-learning* section represents the features of m-learning that relate strongly to social constructivism, and there has already been much written about this (Manuguerra and Petocz, 2011 Wang and Shen, 2012; Laurillard et al., 2013; Martin and Ertzberger 2013).

The FRAME model intersection relating to *context learning* seems to cover an area that has been less represented in the topic literature. This intersection relates to device usability by the learner, Koole and Ally (2006) state that this part of the model relates the characteristics of mobile devices to cognitive tasks associated with the manipulation (and storage) of information and that these processes affect cognitive load. A thorough search of the topic literature relating to cognitive load and m-learning showed very few studies in this area, which is likely to be due to the fact that smartphones and particularly tablets are a relatively new phenomenon. The apparent lack of research in HCI as a source of cognitive load when using mobile devices as educational learning tools suggested that there might be a contribution to be made to knowledge in this area. It also provided a very good fit with the researcher’s own interests in cognitivism, material design and medical education.

Medical education often requires the learner to be in a learning environment other than the classroom. This situated learning leads to the concept of *situated cognition*, whereby thinking is embedded in the specifics of a particular encounter or context. Young et al. (2014), indicate that in a clinical setting there may be participants other than the learner, perhaps a patient, or other staff members. There may also be very different learning environments such as the accident and emergency department, the operating theatre or the ward. From the learner’s point of view any increase in the number of environmental elements involved in learning will also increase cognitive load and present a barrier to successful schemata formation and therefore to learning (Sweller, van Merriënboer and Paas, 1998; Piaget, 2001). Using a cognitive strategy in learning is somewhat in

contradiction to the constructivist approach favoured by modern educationalists. CLT relates strongly to instructional design, and the didactic method of teaching, whereas constructivism emphasises the conceptualisation of knowledge through social interaction. However, in medical-education and particularly computer-mediated learning, there is still a need for instructional design (Young et al., 2014). Learning anatomy often requires rote-learning. Learners are often required to memorise long lists of anatomical structures in the correct order, such as the names of the cranial nerves or the names of the bones of the wrist. There are also sets of skills that must be learnt that are firmly in the psychomotor domain (Bloom, 1956; Dave, 1967; Anderson and Krathwohl, 2000) such as operating an endoscope or conducting an endotracheal intubation. These areas of learning may require instructional design to be considered, and although there may be room for constructivist approaches, cognition plays a significant role.

Mobile devices have been identified by various authors as having the facilities to aid situated cognition and provide a platform for multimedia learning that has been found to encourage engagement (Manuguera and Petocz, 2011; Clark and Luckin, 2012). However, as a piece of computer hardware, mobile devices are also capable of adding to the cognitive load of the user (Hollender et al., 2010). This additional load can be due to many factors including software design, hardware limitations, network issues and distraction by non-task-related events. All of these are well-known in the field of HCI (Hurtienne 2009) and have been studied exhaustively in the field of traditional computing. Mobile devices, on the other hand, do not yet appear to have been fully assessed from educational perspectives such as CLT. Mobile devices offer new physical features that are not found in traditional computers. These include capacitive touch-screens and an array of sensors including global positioning, heat and light sensors, gyroscopes and latterly physiological measurement tools such as heart-rate monitors and pulse oximeters (Deegan and Rothwell, 2010; Deegan, 2015; Martin and Ertzberger, 2013). Many of these features can be used in education and research, and offer a new platform and new strategies for the presentation of learning materials. There is, therefore, a potential for assessment of these devices as learning tools and because they are human-computer interfaces, a cognitive approach is favoured. This is not merely because cognitivism is aligned with the tradition of HCI, but also because effective learning requires cognition (van Merriënboer and Ayres, 2005).

At the time that this study was first proposed (2012) there was little published research on CLT in m-learning. In the four years since there have been a number of publications, some looking at software, a few looking at hardware, but no studies comparing the differences in cognitive load between m-learning on a touch-screen device and traditional classroom or paper-based learning. This chapter is intended to provide a rationale for the research presented in this thesis by critically evaluating the recent literature that has been published in the field.

This review is not intended to be systematic in the formal sense of the word, but to ensure that the review was conducted in a rigorous way, a systematic approach was used to select the papers for review, and also to identify the key themes relating to m-learning and CLT. A preliminary scoping search of the literature was conducted using key search terms to include *multimedia, touch-screen devices, m-learning, learning technology, cognitive load theory, e-learning, m-learning, cognitivism, distance learning and computer-based learning*. The resulting range of publications was then narrowed down using Boolean search terms AND, OR and NOT to obtain titles that were more specific to the topic of m-learning on portable computing devices such as smartphones and tablet computers. Advanced search functions were also employed to narrow the search to recent publications (unless relevant from a historical perspective) and to ensure that key authors and key sources were represented in the search. The literature extracted included research papers, journal articles, policy documents, conference proceedings, press and media releases, webinar content and some grey literature. An initial reading was conducted from which a conceptual map (Figure 2-2) was created to help identify key themes that emerged from the literature.

This diagram helped in identifying contemporary issues, looking for distinctions and connections and identifying where there was scope for informing professional practice. The diagram also identified some key questions that were required to increase understanding and knowledge of the subject area. These themes included: identifying the origins and definitions of the topic, querying the epistemological and ontological grounds behind the area of interest, looking at major issues relating to the topic and investigating

how knowledge on the topic is structured and organised. Many of the papers were not specifically related to m-learning or cognitive load and were therefore not focused enough for inclusion in this review chapter.

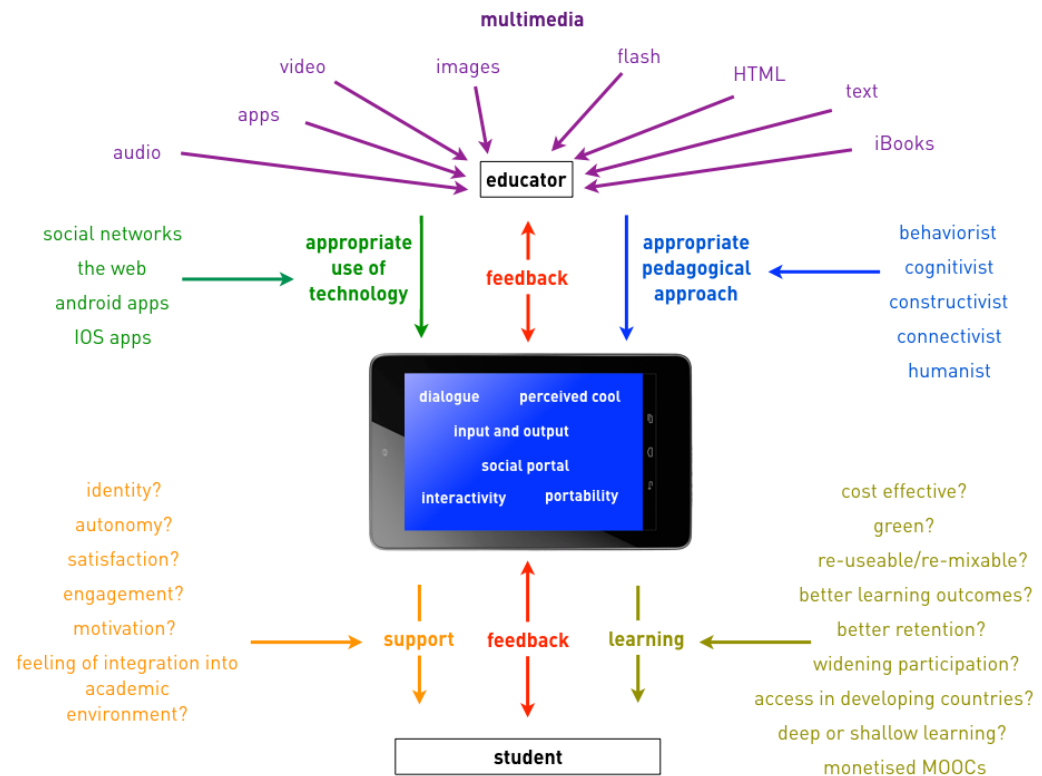


Figure 2-2: A conceptual map providing a technical and pedagogical overview of m-learning

Having performed a broad search, the most relevant texts were then revisited in order to critically evaluate the contemporary issues in m-learning related to cognition and cognitive load. Further searches were carried out using terms that were more focussed on m-learning and CLT. Emphasis was given to recent papers and conference proceedings (since 2010) were included as these were more likely to include the use of tablet devices, and would reflect the current state-of-the-art in terms of device functionality (such as screen resolution, choice of screen size and processor speed). This decision was later

reinforced by a finding in a paper by Sung, Chang and Yang (2015) who presented a histogram of devices used in m-learning (in languages) that shows a sevenfold increase in the use of mobile devices in the year 2010.

All of the papers were skim-read, and the abstracts scrutinised for relevance to the research question and for quality assessment. The papers were then checked against the current *SCImago Journal and Country Ranking* system (SJR). This system of ranking employs the same algorithms used to determine journal impact factor and measures the scientific influence of an average paper from a particular journal based on citations per document over a two-year period. In view of the recency of m-learning, it was noted that the number of publications on this topic is relatively limited in comparison with other types of education theory, particularly papers that also include CLT. To maintain a broad enough range of sources for this review and avoid publication-bias, all papers from ranked journals were considered for inclusion. Two papers were excluded as they were works-in-progress and did not offer any results. Papers from unranked sources were checked for quality and all were excluded. Limitations of these papers included misunderstandings about the definition of m-learning, non-statistically significant findings, spelling errors and typographic errors, inappropriate sampling (such as the use of male-only participants in papers from Saudi Arabia) and out-of-date references, some of which suggested recency bias (whereby trends from a particular time period are predicted to continue into the future). This type of bias is particularly applicable to m-learning due to the rapidly-changing nature of the underlying technology. For example, a paper from 2006 noted network-speed as a barrier to m-learning, but was quoted in a paper from 2015 by which time mobile networks had improved in speed with the introduction of the fourth-generation (4G) mobile communication technology standards. Having determined the studies of high enough quality for inclusion, key points from each paper were entered into an Excel spreadsheet (Microsoft Inc.) to give a structured view of the concepts covered. Some papers did not specifically include CLT, but used methods such as pre and post-testing to determine the effectiveness of m-learning in various contexts. These were included in the review, as pre/post-testing is an indirect measure of

cognitive load.

2.2. COGNITIVE LOAD THEORY IN INSTRUCTIONAL DESIGN

CLT originated in the 1980s but has its origins in the cognitive psychology of George Miller, author of the seminal paper *The Magical Number Seven Plus or Minus Two* (1956). In this paper, Miller was one of the first psychologists to identify limitations in the human working memory, noting that the average person could only store or process around seven items of information at a time. Cowan (2001) suggested that this capacity could be as low as four items. Recent experiments using fMRI have supported both of these early studies, revealing that the working memory capacity in adults is between four and seven items (Arsalidou, et al., 2013). In the field of educational psychology, one early model (Atkinson and Shiffrin, 1968) identifies the working memory as the short-term memory used for processing cognitive tasks before the information is either forgotten or stored in the long-term memory for later retrieval. This model was expanded by Baddeley and Hitch in 1974 who divided the working memory into separate “slave systems”.

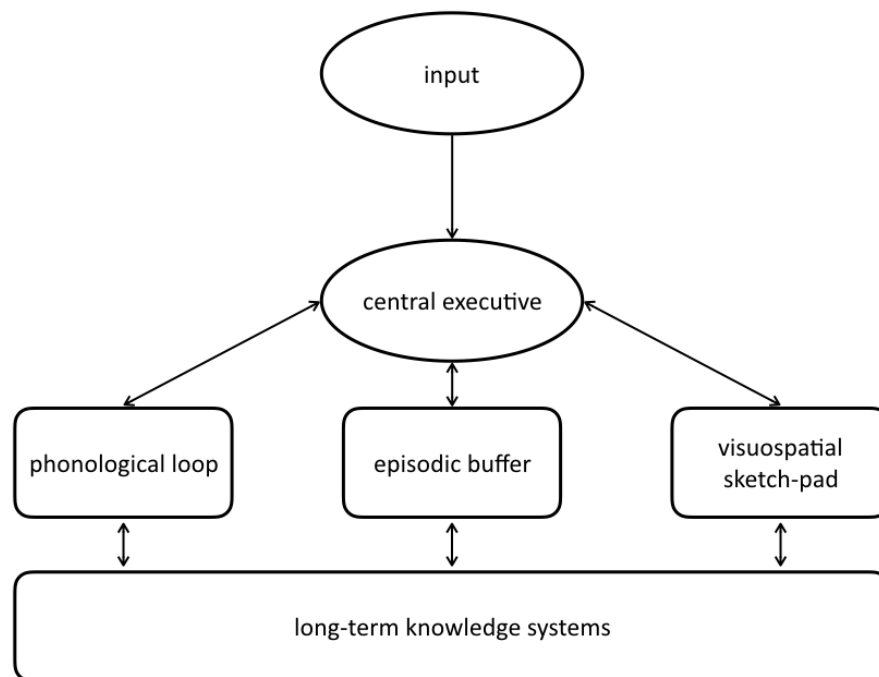


Figure 2-3: model of working memory proposed by Baddeley and Hitch (2000)

Two of these systems are domain-specific, in that they relate to the processing of auditory and visual information separately. The systems are known as the *phonological*

loop and the *visuospatial sketch-pad*, and they are key to the reduction in cognitive load by a principle known as the modality effect which is described on page 66. The *episodic buffer* can temporarily store information from both audio and visual modes and is thought to “chunk” the information. Chunking is necessary for the storage of information in long-term memory; these chunks form schemata, defined as organised thoughts each containing multiple elements of information.

CLT is defined as “a framework of instructional design principles based on the characteristics and relations between the structures that constitute human cognitive architecture, particularly working memory and long-term memory” (Wong, et al., 2012 p.449). Schemata are integral to CLT because although a single schema may have originally been constructed from of a number of elements of information, when these are combined, a schema is handled by working memory as a single element. This may affect cognitive load in some learners as it influences a factor known as element interactivity which is explained in section 2.2.1.

CLT has become an increasingly important factor in the field of instructional design (Sweller, 1994; Mayer, 2014). This is due to the fact that there is now a large body of research evidence supporting the assertion that certain instructional procedures are more effective when based on CLT. The underlying principle is that the information presented to students and the activities undertaken in learning should recognise limitations in the working memory of the learner. Chandler (2004) was one of the first researchers to realise that CLT plays a key role in the construction of e-learning tools and that these tools should be designed with the cognitive load of learners in mind. CLT identifies three distinct types - *extraneous*, *intrinsic* and *germane* although some recent thinking incorporates the latter two types (De Jong, 2009; Wong, et al., 2012). These are explained in the following sections.

2.2.1. Intrinsic Cognitive Load

Intrinsic cognitive load (ICL) relates to the intrinsic nature of a learning task and largely depends upon the number of task elements that must be simultaneously processed by the working memory of the learner. This concept is known as element interactivity, Sweller (1994, p.58) defines interacting elements as “those that must be processed simultaneously in the working memory because they are logically related”. As previously stated, schemata-formation relies on the integration of chunks of information, and before the schema can be formed the chunks must be stored (temporarily) as individual elements in working memory – namely, the episodic buffer described by Baddeley (2000). If a learning task requires the learner to process one element at a time, element interactivity is said to be low and results in a low intrinsic cognitive load. Sweller, Ayres and Kalyuga (2011) provide the example of learning a foreign language, where one can learn the French word for cat (chat) in isolation from the French word for dog (chien). There is no logical connection between the two, and therefore they do not need to be learnt simultaneously. Conversely, if a learner is using algebra to solve an equation (such as $a+b/c=d$), there are many inter-related elements. Understanding the meaning of each element is required, the numbers, the symbols and also their relationships with each other. The equation cannot be solved unless the learner can process many, or all of the elements simultaneously. This is defined as high element interactivity and leads to a higher intrinsic cognitive load. ICL is therefore largely beyond the control of the teacher or student because it will be determined by the nature of the learning task.

2.2.2. Extraneous Cognitive Load

Extraneous cognitive load (ECL) is influenced by the design of the instructional materials and the process of learning (Sweller, 1994; van Merriënboer, 2010). If materials are presented in a way that is not conducive to learning, or contain extraneous content that does not specifically relate to the learning outcomes of the task, mental effort may be invested in these elements unnecessarily. This effort may be at the expense of the mental resources required to integrate the task elements that are germane to learning. ECL is, therefore, a parameter that can be influenced by the teacher/instructor/designer. During a learning task, the working memory will be processing elements relating to both ICL and

ECL (as they are additive). If these two combined sources of load exceed the limited capacity of the working memory (Miller, 1956), cognition will be hampered, and learning will not occur.

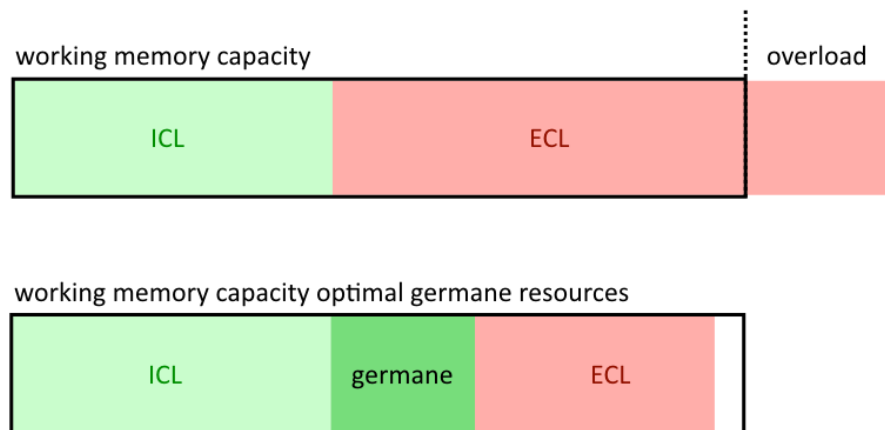


Figure 2-4: (after van Merriënboer and Sweller, 2010) ICL and ECL are additive, if they exceed the capacity of the working memory, germane resources cannot be employed in the formation of schemata.

Under certain circumstances, the presence of many active schemata (for example as may be found in an expert in a particular subject) can lead to extraneous cognitive load if the learning-materials are pitched at the wrong level. The reason for this is that a novice (having no formed schemata about a subject) will require a high degree of instructional guidance because they have no pre-existing knowledge about the subject in question. An expert, on the other hand, may have many pre-formed schemata about the subject and consequently will be guided by this existing knowledge. If a learning activity contains instructional guidance that is already known to the expert, their working memory will be required to process the existing schemata-based elements and also the redundant instructional information. This may consume sufficient working memory resources to cause cognitive overload and is known as the *Expertise Reversal Effect*. This is the primary reason that learning materials must be pitched at a level that is congruent with the academic level of the learner. (Kalyuga, et al., 2003)

2.2.3. Germane Resources (Germane Cognitive Load)

The term *germane resources* refers to the working memory of the learner, and represents the resources that remain available to deal with the intrinsic cognitive load created by the learning task, construct schemata and allow learning. It therefore has an inverse relationship with ECL, because as this factor increases there are fewer resources left available for the processing of information into long-term memory and the creation of schemata (van Merriënboer, 2010).

Germane resources are also often referred to as *germane cognitive load*, but this has been revised by some authors following a reappraisal of CLT by De Jong in 2009. De Jong suggested that germane load might be considered to occupy a different ontological category to intrinsic cognitive load, specifically that ICL (and ECL) are determined by the materials, and GCL relates to the cognitive (working memory) resources of the learner. By this definition, De Jong argues that it would not be possible to add them together (i.e. combine them as one and the same category).

Cognition (and therefore learning) is impaired if the sum of the intrinsic and extrinsic cognitive load occupies working memory to the extent that it prevents the processing of information by the germane resources of the learner.

2.3. CLT AND LEARNING

Sweller (1989, 1994) has demonstrated that a reduction in cognitive load increases learning efficiency. Mayer (2009 p.22) reinforces this view stating that “meaningful learning depends on the learner’s cognitive activity during learning and that well-designed multimedia instructional messages can promote active cognitive processing in learners”. Mayer puts this idea into context describing a typical strategy for designing a multimedia presentation. It requires content; it should be aesthetically pleasing, it should take advantage of technological resources, but if the design is based on intuition rather than scientific research, the tool that has been developed may be incompatible with the way people learn. In addition to material-design considerations, there are a number of published works indicating that the use of increasingly complex technological interfaces may also have a negative effect on cognitive load (Hollender et al., 2010; Schmidt-Weigand, 2011). It is, therefore, important in multimedia learning (and m-learning) to understand the material-design factors that can affect extraneous cognitive load in the learner. (Sweller, van Merriënboer and Paas, 1998). By understanding these effects, it may be possible to design learning materials that off-set the cognitive load placed on the learner by the technological interface.

The commonly cited strategies are:

2.3.1. Goal-Free Effect

This is where a learning activity is not given a specific goal, but learners are asked to find out what they can about a particular problem or question. This approach to learning is thought to reduce cognitive load because, in a goal-free environment, the learner is encouraged to look at one issue at a time rather than having working memory overloaded by the high element interactivity that is known to be associated with means-end analysis. (Sweller, Ayres and Kalyuga, 2011) In other words, if the learners are using working memory resources to focus on the goal there will be fewer resources remaining available for learning the stages of the task.

2.3.2. Worked Example Effect

This effect is seen when novice learners are presented with examples of solved problems. Cognitive load is reduced because comprehending the stages required to solve a problem reduces the high element interactivity that would be associated with having to work out the solution without guidance (Kalyuga, et al., 2001). This resonates with Vygotsky's Zone of Proximal Development (Vygotsky, 1986, p.86) namely "the distance between the actual developmental level as determined by independent problem-solving and the level of potential development as determined through problem-solving under adult guidance, or in collaboration with more capable peers".

2.3.3. Split-Attention Effect

This effect is determined by how learning materials are graphically presented to the learner. Presenting sets of information in a closely integrated way – such as a diagram with overlaid labels prevents the learner from having to divide their attention between, and then integrate, multiple sources of spatially (or temporally) separate information (Mayer, 2009).

2.3.4. Modality Effect

In this scenario, two elements of similar information are provided via different sensory modalities – such as images and audio (Ginns, 2005). The theory has its roots in what is known as Paivio's *Dual-Coding Theory* of 1971. Paivio (1990) postulated that the reason mental images can aid learning is that visual and verbal information can be processed simultaneously through different channels by the brain. The modality effect also holds that audio information is encoded by the brain in a separate stream to visual information and that the acoustic code persists for a longer time in working memory. The effect relies on the fact that auditory and visual information have independent processing buffers in working memory as identified by Baddeley and Hitch (1974) and is related to a construct known as the *separate streams hypothesis* (Penney, 1989).

2.3.5. Redundancy Effect

This relates to materials where information is unnecessarily duplicated, for example, a self-explanatory diagram having a superfluous accompanying description or a narration of text that is already presented visually. This effect also applies to live classroom presentation when a presenter reads the text verbatim from PowerPoint slides. In this scenario, cognitive processing is wasted in attempting to co-ordinate the two elements (Mayer, 2009; Sweller, 1989). Cognitive load may be reduced by eliminating the unnecessary element. In the classroom or in an m-learning application the narration should literally speak for itself and may be accompanied by graphical content or diagrams rather than text.

All of these techniques may be used in the design of m-learning materials, and there is also a body of evidence to suggest that there are certain features that are unique to m-learning that may also decrease extraneous cognitive load. Pachler, Bachmair and Cook (2010) state that mobile devices can store information for rapid retrieval and therefore free the user from having to commit such information to memory – the concept known as *distributed cognition* described in Chapter 01. This might include the use of a notepad feature, calendar or voice memos.

Touch-screen devices also encourage the learner to interact in real time with computer-generated three-dimensional objects or control the speed of video presentations and animations to match the processing power of their own working memory (Baddeley, 1992; Lowe, 2004; Price and Rogers, 2004; Schwan and Riempp, 2004; Boucheix and Guignard, 2005). This user-controlled animation has also been found to be effective when students have a low spatial awareness. An example of this might be that they are less able to imagine how the mechanical components in a static diagram would behave when in motion (Hegarty in Mayer, 2005).

Mobile devices are ideally configured to present multimedia materials (images, video, text and audio) as they are typically equipped with high-definition screens and high-quality digital audio. Accordingly, they can take advantage of the separate streams hypothesis associated with the modality effect. Penney (1989) looked at how separate

streams of stimuli might affect the way that information is processed in the working memory. An example of a separate streams approach could be a diagram that is presented with an accompanying auditory explanation rather than textual labels. Penney's research expanded on Paivio's dual-coding theory (Paivio, 1990) and Baddeley and Hitch's (2000) models of working memory and discovered that information presented as sound is encoded by the brain using an acoustic code and a phonological code. The former relating to the processing of sound, the latter relating to a process by which the learner silently articulates visual information such as words. Penney's hypothesis was that the auditory code persists in a way that boosts the recall of auditory information in comparison to visual information in working memory tasks.

This concept was reinvestigated in 2003 by Leahy, Chandler and Sweller at the University of New South Wales and again in 2005 by Ginns at the University of Sydney. In this major study Ginns used meta-analysis to investigate the modality effect, Ginns defines this effect as "the educational practice of presenting to-be-learned graphical material visually, and related textual information through an auditory mode" (Ginns, 2005, p.313). He tested three hypotheses, all of which relate to multimedia learning, specifically:

- The presentation of materials as a combination of multimedia would be more effective for learning than presentation in a visual format only.
- The strength of the above effect is moderated by the pace at which the materials are delivered, for example, it is less important when a learner can choose their own pace allowing them to learn at their leisure with no cognitive overload of working memory.
- The strength of the above effect is also moderated by the degree of element interactivity of the learning materials.

The studies chosen for inclusion in Ginns' meta-analysis represented the combined performance of 1887 students and were extracted from published articles, book chapters, conference papers or PhD theses. Only quantitative studies were included where statistical evidence could be combined. The results supported all the hypotheses put

forward. It was confirmed that when verbal and visual information (images) were presented together, the students out-performed those who were provided with visual material only. Presenting the stimuli simultaneously is known as *temporal contiguity* and has its roots in the field of behaviourism where it is necessary for the effective application of both classic and operant conditioning (Skinner, 1988). The effect was reduced however when the “visual material only” students were allowed to study at their own pace. They had the time to absorb an equivalent amount of information using only a visual pathway. Ginns concludes that learners do not always have the luxury of unlimited time-frames, they are usually required to work to deadlines and that the use of correctly applied multimedia in instructional design should result in a more efficient use of cognitive resources in the learner.

In addition to auditory and visual components, the responsive screens of Mobile devices afford another sensory channel to and from the learner, namely touch, and haptic feedback; whereby the device provides audio feedback and vibrates (or exerts a force) simulating a response to touch (MacKenzie and Oniszczak, 1998). The effect of tactile interaction on the modality effect seems to be an under-researched area. In 2005 Conway and Christiansen stated in a related research project that “few researchers have conducted rigorous comparisons across sensory modalities; in particular, the sense of touch has been virtually ignored” (p.24). It would, therefore, be interesting to discover if the tactile component of touch-screen devices has an effect on learning from a cognitive load or modality-effect viewpoint.

2.4. CRITICISMS OF COGNITIVE LOAD THEORY

The principal reason for using cognitive load theory to underpin this research was because it offers a well-recognised structure on which to base learning material design, particularly in areas that are technically challenging (Hollender et al., 2010), and in medical education (van Merriënboer and Sweller 2010). It is, however, important to recognise the potential shortcomings of the model, and take these into account when designing the data collection method and analysing any conclusions drawn from the research. Despite its widespread use in psychological research, CLT and the nature of cognitive load has been a disputed subject over the last 30 years. Moreno (2010) explains that there are many conceptual, ontological and methodological limitations that may

reduce the scientific rigour of the theory. Whelan (2007) states that task complexity may blur the boundaries between intrinsic, extrinsic and extraneous cognitive load and (as stated on p. 64) de Jong (2009) questions the additivity of these elements given that germane resources appear to occupy a different ontological category to intrinsic and extrinsic cognitive load. Ayers and Paas (2012) have also reappraised CLT and defended the use of self-rating scales over dual-task methods of cognitive load measurement (these data collection methods are fully evaluated in chapter 03).

2.4.1. Conceptual Criticisms of CLT

One of the main conceptual criticisms of the model relates to the distinctiveness between the three constructs of intrinsic, extrinsic and germane load as proposed by Sweller (1994). A commonly cited premise concerning intrinsic cognitive load is that it is fixed and cannot be changed by the instructor (Ayers 2004). Schnotz and Kürschner (2007) assert that intrinsic cognitive load can only be considered a fixed element in the context of a particular learning task and at a particular level of expertise. They give the example of a document written in legal jargon that would be difficult to understand by a lay person (increasing extraneous cognitive load) yet would be expected to be learnt and understood by law students and could be considered to be load that is intrinsic to the learning task. In a similar vein, Leppink and van den Heuvel (2015) state that there has been much indecision over whether to favour a three-factor (triarchic) model of cognitive load, or to return to an earlier two-factor model in which germane load is considered to be a subdivision of intrinsic cognitive load. Moreno (2010) also asserts that factors that appear to induce ECL may also induce GCL and vice-versa. In summary, the differences between the types of cognitive load may not be as clear-cut as Sweller's 1994 classification suggests.

In defence of the conceptual limitations defined above, some of the criticisms relating to the triad of constructs described by Sweller (1994) appear to relate to somewhat semantic differences. Whether referred to as *germane load* (Sweller 1994), *germane resources* (Kalyuga, 2011) a *subsection of intrinsic load* or a *subjective judgement of learning* (Leppink and van den Heuvel 2015), there is empirically-derived evidence to show that working memory has limitations and all of the above terms are used to

describe resources that are available for learning (Miller, 1956; Baddeley, 1992; Cowan 2001; Brünken, Plass and Leutner, 2003; Wong, Leahy, Marcus and Sweller, 2012; Leppink and van den Heuval, 2015). Whether or not one considers germane resources to be a subdivision of intrinsic cognitive load or not, this issue becomes less relevant in a study such as this, where only a single learning activity/topic was employed. This is because the intrinsic cognitive load will be the same for both groups and any differences in reported task load will therefore not be due to this construct. The idea put forward by Moreno (2010) that factors may affect both ECL and GCL seems, by definition, to be self-contradictory. Cognitive load that is expended in the formation of schemata, cannot be defined as extraneous by Sweller's (1994) definition.

2.4.2. Methodological Criticisms of CLT

One of the primary limitations of CLT highlighted by de Jong (2009), Brünken and Seufert (2009) and Brünken, Plass and Leutner (2003) relates to whether it can be accurately measured. Early studies relied on indirect methods such as counting error rates, or time-on-task, however more recent research has adopted a self-reported task load measure first developed by Paas in 1992. This method only reports perceived mental effort and is considered by some authors to be inferior to direct objective measures of cognitive load (Brünken, Plaas and Leutner, 2003). Having a single factor (mental effort) under test precludes any evaluation of internal consistency that could otherwise be achieved by examining correlations between different items that propose to measure the same construct. Furthermore, mental effort is considered to be only one of many contributors to cognitive load. Van Gog and Paas (2008) have argued that the Paas scale has been modified by different researchers in the field and that the resulting lack of consistency may have resulted in constructs other than cognitive load being measured. A final concern relating to self-reported effort is that subjective scales such as this are not as robust as dual-task methodology. Dual-task methods (covered in more depth on page 109) require the participant to undertake a secondary task such as finger tapping, the performance of which provides an objective measure of the load being placed on the participant by the primary task.

From a methodological standpoint, there are strategies that can be employed to offset the limitations identified above. Despite the reservations expressed by Brünken, Plass and Leutner (2003) de Jong (2009) and Brünken and Seufert (2009) measuring cognitive load can be achieved using a variety of data collection methods. These are fully evaluated from page 105. Issues relating to consistency of data collection can be addressed by using a standardised method such as the NASA TLX which is fully appraised on page 113. To ensure consistency with previous studies the design of the NASA TLX was not adapted or modified from the original tool described by Hart and Staveland (1988). This tool is multi-dimensional and measures more than the single construct of mental effort improving internal consistency by allowing correlations to be made between mental and physical demand, temporal demand, performance, frustration and effort. A comprehensive justification for using the NASA TLX as a measure for cognitive load can be found on page 119.

2.4.3. Ontological Criticism of CLT

The main ontological criticism of cognitive load theory also relates to how the three main constructs are defined. In this case, de Jong (2009) raises a question about the nature of germane load. Much of the preceding research relies on the premise that the constructs of cognitive load can be added together (Sweller, 1998), however if intrinsic and germane cognitive load are considered (respectively) to relate to the *material design* and the *cognitive processes* (required to form schemata), they cannot be considered to belong to the same ontological category. This, in turn, means that they cannot be compared or meaningfully added together.

De Jong's (2009) ontological criticism of germane and intrinsic load appears to be open to interpretation, or possibly a misunderstanding of Sweller's definition. Van Merriënboer and Sweller (2010) define intrinsic load as the *working memory load* affected by the intrinsic nature of the learning tasks. If intrinsic cognitive load is considered to be cognitive load generated by the inherent complexity of the learning material rather than referring to the actual material itself (in de Jong's words "an object") there seems to be little ontological distinction between this and the working memory load that is germane

to schemata formation. They can both be classified as types of load placed on working memory, i.e. cognitive processes rather than material design.

The relationship between pre-existing knowledge and ICL is well known. The relationship between cognitive load and expertise (expertise reversal effect) has already been evaluated on page 63. The expertise reversal effect can be reduced by choosing a learning topic about which the participants are unlikely to have previous knowledge (expertise). This is explained in more detail on page 120. It is important to remember, however, that a degree of cognitive load is required to optimise performance in a learning task (Chen, et al., 2016) and this is further covered on page 251.

In this study it was decided to preserve the triarchic model of cognitive load (ICL, GCL and ECL) because despite philosophical objections in the literature (Schnotz and Kürschner, 2007; Leppink and van den Heuvel, 2015) the model appears to be objectively supported by neuroscience and more recent psychological research. Whelan (2007) explains that although traditional measures were unable to accurately distinguish between different types of cognitive load, functional neuroimaging techniques appear to be able to observe these effects in real-time and also locate distinct areas of the neo-cortex that relate to each type of load. ICL has been linked to activity in the dorsolateral prefrontal cortex, GCL in the superior frontal sulcus and intraparietal sulcus and ECL in Broca's area, Wernicke's area and the posterior parietal association cortex. A more in-depth explanation of functional neuroimaging can be found on page 106 together with an explanation as to why it was not suitable for use in this study. Other recent research also supports the triarchic model, including at least one study that distinguishes between GCL and ICL (Dehue and van de Leemput, 2014).

In summary, like any other theoretical model, CLT is open to scrutiny and has been criticised by authors including de Jong (2009) and Moreno (2010). Many of the criticisms appear to be somewhat superficial, may show a misunderstanding of CLT and do not appear to offer a robust alternative. In 2009, De Jong recognised that modern neuroimaging techniques might offer promising new methods for measuring cognitive load yet these techniques have come to support the original triarchic model that de Jong questioned. Historical criticisms of the theory do not suggest that CLT should be discontinued, but should be developed in line with the principles of best practice in science research. This seems to have been the case and the model continues to be

updated and used in numerous experimental studies especially in the area of medical education (Fraser, Ayres and Sweller, 2015). In a recent communication, Sweller (2018) states that “ironically, much of the theory’s impact seems to have occurred in the years following the publication of... papers (criticising the model)”. In the context of my study, CLT provides a useful framework on which to base the choice of data collection method and also evaluate the empirical findings from a conceptual perspective. It is important, however, to recognise the limitations of the theory and to ameliorate the potential pitfalls that these limitations may introduce.

2.5. THE COGNITIVE THEORY OF MULTIMEDIA

Cognitive theory is used in the field of m-learning because if a device or the learning materials are not compatible with the formation of complete schemata, there is likely to be a barrier to learning. Mayer (2009) posits that the rationale behind the use of multimedia is the fact that humans learn more deeply from materials that contain words and images than when learning from words alone. Mayer recognises that the pictorial content of multimedia has evolved to include computer generated imagery and the use of technology to deliver the content, but asserts that simply adding words to pictures does not necessarily improve learning outcomes. Learning materials must be presented in a way that fosters understanding, and there are many features of instructional design that may hamper cognition if used incorrectly or inappropriately. Mayer refers to the split-attention effect, first described by Tarmizi and Sweller (1988) whereby learners have difficulty when graphic elements are not closely integrated with accompanying text. Kalyuga, Chandler and Sweller (1999) concur that graphical elements that are separated by distance (i.e. presented at different locations in a document) or time (presented at different times during a presentation or activity but not concurrently) are difficult to mentally integrate. The reason for this difficulty is that the learner must invest time in searching and recalling certain elements and this will occupy resources in working memory. (Kalyuga, Chandler and Sweller, 1999)

Ayres and Paas (2007) go further, stating that some *single* elements of multimedia, such as animation, may hamper learning if they pose excessive demands upon the working memory of the learner. Animations are often cited anecdotally as being “distracting” when used in teaching materials. But it is arguable whether this is due to the animation itself, or how it is used. Animations can give rise to the split attention effect, but also may violate a concept known as the “apprehension principle” which relates to a measurable reduction in the user comprehension of an animation compared to static images (Tversky, Bauer-Morrison and Bétrancourt, 2002). However, a more recent meta-analysis by Höffler and Leutner (2007) has shown that animations, in themselves, may not be the cause of learner misapprehension. Their study revealed an overall advantage could be measured when dynamic visualisation was used compared to static images. However, this relies on certain conditions, most importantly that video should be representational of the topic to

be learnt. Decorative animations or images that do not relate to a topic, such as animated comic clip-art, offer no advantage over static images and can increase cognitive load if they divert the learner from the message that the lesson is intending to convey. This is known as the *seductive detail* effect, first described by Harp and Mayer in 1998. The mechanism of this effect is that the additional content (detail) seduces the learner away from the content that pertains to the learning topic. The insightful example given by Harp and Mayer relates to an instructional exercise on *the meteorological causes of lightning*. If the seductive details show imagery of damage caused by lightning strikes, the student will be seduced into thinking that the lesson was about “what lightning causes” rather than “the causes of lightning”.

There are, therefore, trade-offs to be made when choosing the types of content to include in teaching materials for m-learning. To achieve an optimal m-learning environment, the developer must clearly take into account the body of evidence relating to cognitive load.

2.6. CLT AND M-LEARNING

The relationship between cognition and e-learning was first formalised over three symposia held in 2004 at the annual meeting of the American Educational Research Association (van Merriënboer and Ayres, 2004). All three events were focussed on e-learning and how the design of e-learning materials could be optimised in the light of recent advances in CLT. The following section critically evaluates current research relating to CLT in the field of m-learning. Having considered the papers that specifically focus on CLT in m-learning it transpired that nearly all of the studies used a quantitative method of data collection and an experimental approach. This reflects the fact that cognitive load is a quantifiable phenomenon (Brünken, Plass and Leutner, 2003). The main themes that emerged from an initial reading of the literature were:

Device usability - software and hardware 17/48

Situated learning 19/48

Material design and mode of presentation. 33/48

Learner engagement/user adoption 15/48

All of these themes relate to cognitive load because they either contribute to it or result from it. The review will now look at these concepts in more detail.

2.6.1. Device Usability

In their paper of 2010, Hollender, et al., examine some of the human-computer-interaction principles that are used in e-learning to reduce cognitive load in the user/learner. Learning activities involve the use of software designed to support learning, but also must take into account the fact that novice users will also be required to learn how to use the hardware and the operating system. M-learning, as a branch of computer-based learning, must therefore integrate user-centred design and learner-centred instructional design principles. Hollender et al. define two models that integrate the principles of HCI and CLT that may be used in the design of complex educational software. The first recognises the importance of reducing extraneous cognitive load by the use of easy-to-learn, easy-to-use software. The second introduces the concepts of CLT to promote germane resources in the formation of schemata. The need for further empirical research into the use of educational software from a CLT approach was identified, but interestingly hardware issues were not identified as being important. Given the fact that the paper was published in 2010, it is unlikely that the authors were aware of the impending proliferation of tablet computing, but smaller touch-screen devices such as iPod touch and smartphones had been available for use in m-learning since 2007.

Lack of data relating to m-learning hardware was also identified in 2013 by Raptis, et al. who noted that previous research tends to focus on user experience, device attractiveness and brand, rather than on the ergonomics relating to physical factors such as screen-size, button position and casing design. Raptis, et al. performed an empirical, experimental study to look at the effect of screen-size on usability. The experimental task was not strictly education-focussed but required the participants to search for

information about various items online. In brief, their main hypotheses stated that increasing screen size would reduce task-completion-time and task-success-rate – a shorter completion-time indicating increased efficiency and a greater success-rate indicating increased effectiveness. The study did not use CLT in its design, however, task-completion-time can be used as an indirect measure of cognitive load (Chen, Epps and Chen, 2011). Three screen-sizes were included in the experiment, specifically 3.5” (inches), 4.3” and 5.3”. To ensure that the only independent variable was screen-size, the other variables were controlled, the devices were chosen to be the same brand (Samsung Inc.), were installed with the same Android software launcher (Google Inc.) and had their processor speeds adjusted to be equal across all three devices. The predominantly male participants were chosen using convenience sampling and were all students aged between 19-30. These demographics are not representative of the average distance learner who tends to be over 30 years of age and is equally likely to be of either gender (Dabbagh, 2007; Aslanian and Clinefelter, 2012).

Data from convenience sampling is not normally considered robust enough to support inferential statistics. However, it was stated that the findings showed no statistically significant effect of screen size on perceived usability or task success rate, but revealed an inverse correlation between screen size and task completion time (i.e. larger screens shortened task completion time). Although task completion time can be an indicator of cognitive load, it is also determined by the nature of the task. In the study by Raptis, et al., the five tasks involved information retrieval from the Internet and three of the tasks required heavy content-scrolling. Scrolling is known to increase task-time (Jones, Buchanan and Thimbleby, 2003), and it is likely that this was the cause of the task-time increase rather than extraneous cognitive load.

The study was limited in that it only included smartphones, not tablets, but a recommendation for further study was made to include devices of larger screen size. This would be to determine whether the task-efficiency correlation continued, and at which point it ceased to be a significant effect. Two of the findings were of relevance to the methodology of the study relating to this thesis regarding software design for the interactive learning group and the choice of devices to be used. Firstly, that the interactive task should employ an architecture that minimised the need for scrolling in order to be a fair comparison (in terms of CLT) to the non-interactive task (which required

participants to look at a non-scrollable labelled photograph). Secondly that a range of screen-sizes should be evaluated and that these should also include tablet devices. Screen size is a recurring theme in device usability, and there have been a small number of other studies that focus on this feature. One early study by Findlater and McGrenere (2008) looked at correlations between screen size and user satisfaction with adaptive web-design, and also a phenomenon known as *awareness*. This is where the design of an interface loses its full feature set on smaller screens and the user may not be aware of certain functionality. The study used a simulation of smaller screens by partitioning a large, high-resolution computer monitor, which is not an ideal method for reasons covered in section 2.6.3. The study did not strictly consider cognitive load but was written from an HCI perspective. It is included here because it supports the findings of Raptis et al. (2013), in that task-time increased on smaller screens. The results of the study also highlighted the need for adaptive technology in user interface design. Adaptive technology is important in material design as it shortens development time. Rather than having to develop separate apps for different platforms, the same single application may be used across many different devices, causing the content to adapt to the screen size in use but still display the content in a user-friendly way. These principles were applied to the software designed for this study as the learning task was required to be used on a range of mobile devices having different screen-sizes. The rationale is covered in more detail in the methodology chapter.

In 2012, Kim and Kim performed an experimental study looking at the correlation between screen size and pre/post-test outcomes in vocabulary learning. In their research, it was intended to compare three screen sizes corresponding to the iPod (Apple Inc.), a typical smartphone and a Kindle tablet (Amazon Inc.). However, there appeared to be a fundamental misunderstanding between *screen size* and *spatial resolution*. These are not the same thing (Trinder, 2005). In a similar manner to the Findlater McGrenere study of 2008, the three different screen sizes were simulated on a PC screen rather than using mobile devices and therefore possibly did not take into account pixel density (i.e. the screen “sizes” were represented by the number of pixels in the display rather than the physical size of the display). This is not a truly accurate representation because a device with a high resolution may not necessarily have a large screen, for example, a Samsung S7 mobile phone has a 5.1-inch screen with a resolution of 3.68 million pixels whereas a 12-

inch laptop computer screen may only have a resolution of one million pixels. Although counterintuitive, the higher pixel density is necessary because the screen of a mobile device is intended to be viewed from a much closer distance than a computer monitor or television screen, and at this distance, the human retina can more easily distinguish the pixel matrix. To account for this, the pixel density must be increased, providing what Apple (Inc.) call a “retina display”. When held at normal viewing distance, the user is likely to perceive a sharper and clearer image than when viewing a larger monitor from a further distance. In the 2012 Kim and Kim study, the “size” of the simulated screens may therefore not have been truly representative of the actual screen sizes if the pixel matrix occupied a larger screen area on the monitor than it would have on the native device. By failing to use devices fitted with a touchscreen, the study lacked the ecological validity that would otherwise have been afforded by such. Furthermore, had this study used authentic devices they would have been installed with different operating systems - iOS (Apple Inc.), and Android (Google Inc.). This may have been considered a confounding variable eliminated by the simulation. However, it is a factor that may affect device usability from an HCI perspective. Despite these limitations, their results concurred with the Raptis study in that the difference in pre/post test scores correlated positively with (simulated) screen-size and taken as an indirect marker of cognitive load this suggests that load is decreased when using a larger screen.

The findings of the Raptis study are supported by contemporary research by Deegan and Rothwell (2010) who also considered usability in mobile devices for m-learning. Their research examined some of the influences that may present a barrier to the implementation of m-learning into mainstream education. The paper structures m-learning into categories and is useful in that it links pedagogical principles with software design and usability. The authors recognise that software developers are not necessarily educators and without an understanding of the end user, or education theory, the resulting app may be overcomplicated, task-inefficient and unsatisfactory to the learner. The categories of m-learning identified by Deegan and Rothwell are: *Learning Management, Supportive, Content-based, Context Based* and *Collaborative*. The first two categories are concerned with the practicalities of course delivery and the nature of the

virtual learning environment. The category concerning collaboration deals with the pedagogical aspects of constructivist learning theory, but the remaining three categories all have implications for cognition. From a content viewpoint, it must be recognised that mobile devices do not offer the same learning experience that one might expect from a classroom activity or computer-based learning. In a classroom or when using a PC, the learner is typically in a relatively quiet setting and has a set amount of time to undertake a learning activity. This is not the case in m-learning where the learner can be in a situation with multiple distractions and is typically using a device having a smaller screen-size than that found on a desktop computer (Terras and Ramsey, 2012).

Screen size has also been identified as a factor that may pose a barrier to m-learning adoption (Liu, Li and Carlsson, 2010). However, since 2010, this issue has now been largely addressed by the introduction of tablet computing, and the more recent trend of “phablet” devices. Phablet is a term used to describe a mobile phone having a large screen size (such as the Galaxy Note series (Samsung Inc.) and the iPhone 6 (Apple Inc.)) or hybrid tablet computers (such as the Galaxy S2 (Samsung Inc.)) that permit mobile network connectivity via a SIM card and a built-in phone. These devices typically offer the benefits of *always-on* connectivity to the web, and a larger screen.

Content must, therefore, be designed with the device ergonomics in mind, and contextual elements relating to situated learning can be exploited. These include the ability of the device to locate its position through GPS and measure environmental factors such as sound, heat, light, and acceleration (Martin and Ertzberger, 2013). The Situated Learning that is afforded by mobile devices allows a degree of authenticity that is not found in the classroom and is founded on the principles of *situated cognition* whereby learning occurs when the learner is allowed to construct knowledge while immersed in an environment that relates to the intended learning outcomes. This is examined in the following section.

2.6.2. Situated Learning

Situated learning relates to the authenticity of context, and the learners’ interpretation of it. From a CLT perspective, situated learning aids cognition by embedding problem-solving and knowledge acquisition into real-life settings. Martin and Ertzberger (2013) explain

that the *here and now* framework associated with m-learning refers to the portability of the devices permitting learners to undertake activities “in the field” and focus on the context. The device supports the learner by offering a two-way conduit for information transfer. Learners can access support materials from the device, but can also make field-notes, take photographs and record sounds that can be instantaneously shared with other users. Wang and Shen (2012) elaborate on this concept, explaining that m-learners can use their devices to gather knowledge that can then be used to make decisions and solve problems in real-life contexts. Wang and Shen make the distinction between “mobile computers” which they define as laptops, and “network centred devices” which they define as smartphones. Five years later, this distinction is no longer so clear-cut. By their own description of the devices the authors identify a mobile computer as a machine that can retain full functionality without internet access, having a large hard drive and a longer battery life than hand-held devices. However it can be argued that in the latest generation of smartphones and tablets, processor power, external storage and battery life can match or even exceed that of some laptops. The iPad Pro (Apple Inc. 2016) being a typical example. These enhanced features make mobile devices an optimal tool for situated learning.

Others argue that there are drawbacks to situated learning from a cognitive load perspective. Terras and Ramsey (2012) discovered that the environment can cause a distraction to the m-learner under certain circumstances. As a learner moves from context to context, environmental factors such as noise, lighting, visibility and comfort levels will change. All of these factors present stimuli that may distract or disengage the learner from the learning process. The devices also encourage multi-tasking, such as switching from a document to an app, or to a browser. HCI studies show that task-switching is also known to impair cognition (Monsell, S., 2003; Edmondson and Beale, 2007), and it is likely that these are the types of activities a learner would be undertaking to support situated learning.

Sølvberg and Rismark (2012) noted that mobile-learners often study at home where they can use their wireless network to stream content such as lectures. There is an assumption that this environment may create a degree of distraction due to background noise, music or television, and that this type of “multitasking” may affect cognitive processing.

However, this may not always be the case. Lee, Lin and Robertson, (2009) conducted research on the effects of background distractions on knowledge acquisition in relation to cognitive-load. Their study used an experimental design involving 130 predominantly female participants (mean age of 23.9 years) who were randomly assigned into two out of three experimental conditions. The participants were required to undertake a learning activity where they were required to read three articles and answer multiple choice questions based on them. The three experimental conditions were;

Silence, where the participants undertook the activity without distraction, *Background*, where the participants were also required to undertake a second task, namely watching a video but were told that they could ignore it if they wished (although these participants were also tested on the content of the video).

Test, where the participants were also required to watch the video, but were told that they would be tested on the content of the video in addition to the multiple-choice questions.

The results showed that there was no significant difference in the test scores between the students who undertook the activity in silence and those who took the test with the video in the background, however, there was a significant difference between the silence group and the test group. This suggests that the presence of background distractions have little effect on cognitive load when the learner is not required to invest cognitive resources in both the learning activity and the background activity, but that when required to undertake more than one task at a time, less information is retained. This is in full agreement with Sweller, Merriënboer and Paas's (1989) theory of cognitive load, because introducing more elements into the learning activity will require the learner to store more of these elements in working memory. If a task with high element interactivity (such as a complex learning activity) is combined with a task having low element interactivity (such as background music) there may be enough germane resources left over to undertake the task successfully, but if the task is combined with another high-element interactivity task, the extraneous cognitive load will hinder learning. Despite the limitations of situated learning, Martin and Ertzberger (2013) also note that because of the ubiquity of m-learning and the volume of interaction with this user-environment, students tend to be very engaged with the learning process. This

engagement may include interaction with other students and professionals through social networks, but also interaction with learning materials appropriate to the context. In an experimental study, the authors looked at the effects of m-learning on student achievement and attitude by comparing an activity undertaken on an iPad (Apple Inc.), an iPod (Apple Inc.) and a desktop computer. The study looked at the use of mobile devices to augment a visit to an art installation. The sample consisted of 109 predominantly female art students aged between 18-22. An educational software app was designed to provide information relating to paintings hanging in a gallery and was designed to display on PCs and tablets. The participants were required to take a pre-test on the subject in question and then were randomly divided into study groups. The computer-based learners were required to view the paintings and then return to the classroom to undertake the e-learning activity. The m-learners were required to use their devices (iPod or iPad) while inside the gallery, to access the e-learning information in context to the paintings they were viewing. The study does not measure cognitive load directly, however, the use of pre and post-testing can give an indirect assessment of this variable. The authors had anticipated that the m-learners would outperform the computer-based learners in the assessment, but interestingly this was not the case.

The results revealed that there was a statistically significant increased performance level in the computer-based learners as measured by the pre/post-testing. This was in spite of the fact that the m-learners were found to be excited and engaged by the use of the technology - factors that may decrease cognitive load. When broken down the results showed that the significant difference was not between CBL and iPod users, but between the CBL and iPad users. This is unusual because it could be assumed that the iPad screen size would offer a user experience that is more closely matched to a desktop computer than a 4in iPod screen. This result was not fully explained in the study.

From a CLT perspective, there are reasons that may serve to explain the unexpected result. Martin and Ertzberger cited user-distraction as a possible cause because user-attitude data collected as part of the experiment showed that the CBL learners were less distracted than the m-learners. This might have been the case, but as stated earlier, there is some controversy in the literature as to whether distraction always causes an increase in cognitive load. Sweller (2011) and Young, et al. (2014) state that external distractions

that are not related to a learning task can impose extraneous cognitive load, however, in a more recent experimental study, Deegan (2015) ascertained that learners are capable of distinguishing between different sources of cognitive load simultaneously – including distraction. In agreement with the Lee, Lin and Robertson study (2009) their experimental findings showed that external distraction did not affect the performance of the primary task. Another explanation for increasing cognitive load in the Martin and Ertzberger experiment is that the m-learners were accessing verbal and visual (pictorial) information simultaneously. This is also debatable, as previously stated, there is a known effect when verbal and *audio* information are presented together, known as the redundancy effect, and this may also apply when graphic and (unnecessary) textual information are provided simultaneously (Sweller, 1989; Mayer, 2009). The redundancy effect causes an increase in extraneous cognitive load because the learner is having to use working memory resources to assimilate redundant information that does not add to the acquisition of knowledge, however, this does not appear to be the case in this instance. The verbal information was being provided as an adjunct to the pictorial information provided by the paintings and should have provided contextual augmentation. It was not redundant information. Furthermore, according to Paivio's dual coding theory (Paivio, 1990), knowledge acquisition should be enhanced when verbal and pictorial information are provided simultaneously. This leaves one most likely explanation - that the split-attention effect was responsible (Ayres and Sweller, 2005). This is where the learner has to conduct searches between two sources of information and mentally integrate them. If the m-learners were having to split their attention between the paintings, and searching the text contained in the app, it might have adversely increased their extraneous cognitive load. The authors correctly identified that these results are in contradiction to Mayer's (2009) contiguity principle. The CBL group were accessing the information after the event of viewing the paintings, and temporally disconnected elements are thought to be more difficult to mentally assimilate.

In a similar study to the Martin and Ertzberger paper, Chu (2014) used an experimental approach to assess the effects of m-learning on a field-trip but used a more CLT-focussed technique. The participants were two groups of school children assigned to two groups. The control group ($n=31$) were required to learn about indigenous Taiwanese culture on

the field-trip supported by human guidance. The experimental group ($n=33$) were required to learn the same topic but with a Personal Digital Assistant preloaded with digital learning materials relating to the context of the study (a traditional Chin-an temple). This is an unusual choice of device for a m-learning study as PDAs are now obsolete. They are not configured for data input in the same way as modern touch-screen devices and have smaller, lower-resolution screens (Kukulska-Hulme and Traxler, 2005; Smørðal and Gregory, 2005). The market share for these devices is now so small as to be negligible, less than 1% of the world market compared to smartphones at 68% (Shirer, 2015).

The learning materials were designed to augment the learning environment using text, images and quizzes that required the learners to answer questions about the temple features. The cognitive load of the participants was measured using a self-reporting (indirect) tool (Sweller, van Merriënboer and Paas, 1998), and pre/post-testing was also employed.

As in the Martin- Ertzberger study, the experimental group was expected to perform better than the control, but did not. The students using the mobile devices demonstrated statistically lower learning achievement in the post-test. The cognitive load measurements were checked to see if there was a correlation between these and the post-test scores. Two dimensions of cognitive load were assessed, mental load and mental effort. Mental load is a task-based dimension, it is defined as the load imposed upon the learner by the inherent task elements and is not amenable to manipulation by the instructional designer. Mental effort can be defined as the cognitive capacity allocated to the task demands (i.e. learning) (Sweller, Ayres and Kalyuga, 1994). Chu discovered that there was no significant difference in mental effort reported between the two groups, but a significantly higher mental load was reported by the m-learners. This was interpreted as having been caused by the fact that the m-learners were having to answer many questions on their devices during the learning activity, and were therefore required to perform a lot of information-seeking during the task.

Chu attempts to rationalise the increase in mental effort, stating that the m-learning environment requires the learners to spend more time comparing real-world (contextual) information with digital resources than they would be required to spend in a purely web-

based environment. This does not seem to provide a satisfactory answer in this case because the control group were not in a digital environment – they were in a real-world environment. Furthermore, no indication is given as to the number of formative questions asked of the control group. However, assuming that all other factors remained the same, it would appear that the use of the mobile devices increased cognitive load in the experimental group, and this is reflected in the post-test scores. Given the findings of the Martin and Ertzberger study, and the similarity of the two approaches, it would be reasonable to hypothesise that the split-attention effect is likely to have played a role in the effect.

Situated learning can also be examined from a qualitative viewpoint, in a 2012 study, Sølvsberg and Rismark gathered rich, descriptive data from a group of Norwegian students relating to the use of m-learning across three different learning spaces. The students were divided into three groups and were required to access streaming lectures in three settings, group 1 watched in the classroom, group 2 in a location on campus and group 3 from off-campus locations. When given a choice most students preferred not to attend the classroom lectures, but this choice was determined by factors such as whether they wanted to ask questions about a particular topic and whether the topic matched their general interests. One important observation noted that attending the class gave the participants a feeling of belonging to a social community, which may be difficult to achieve in a m-learning environment where the students are remote from each other.

The group of participants who accessed the materials on mobile devices but did so on-campus, benefitted from the social aspects of communal learning but with the flexibility to choose their own time to learn. The participants set up impromptu learning spaces and in the words of one student found themselves to be “more concentrated... in class, there are lots of other distractions” (Sølvsberg and Rismark, 2012, p.28). This point is interesting in the light of studies evaluated in the previous section where the mobile environment was hypothesised to be more distracting than the classroom.

The students who undertook the learning off-campus typically took advantage of using their own wireless networks and studied in their homes. The main theme that emerged from this group was that the learning became fragmented – interspersed and sometimes combined with domestic activities. This often led to materials such as lecture videos being

only part-watched and it was thought that the resulting learning lacked depth. From a m-learning material design point of view, this is useful information because it supports the theory that materials should be broken down into smaller chunks. Chunking is a well-known technique in the field of instructional design as it decreases cognitive load. Information that is separated into smaller chunks of conceptually-related material helps to create schemata because, in essence, they are the same thing. Cooper (1998) gives some interesting examples of this technique – such as how a telephone number is easier to remember when split into chunks of digits rather than being presented as a single 11-digit number or how a shopping list is easier to remember if the items are chunked into conceptually similar sections:

Apples, Pears, Oranges
Bread, Butter, Jam
Pens, Paper, Envelopes

rather than

oranges, paper, jam, envelopes, bread, pears, pens, apples, butter.

For mobile developers, this gives a clue that larger subjects should be broken down into smaller topics and each topic presented as short easily-digestible chunks that can be browsed in an informal learning environment where long free periods of time may be few and far between. The students on the off-campus group in the 2012 Sølberg and Rismark study gave a compelling example of this when they explained that the video lectures were never accessed on public transport journeys (which would normally make very useful free time for learning) because the journey home was shorter than the length of the video that they were required to watch.

2.6.3. Material design and mode of presentation

There is much debate over material-design for m-learning. Wang and Shen (2012) advise

in their paper on design principles for m-learning, that materials should be created for general e-learning (for example to be used on desktop computers) and adapted for mobile devices. This is in direct contradiction to the design principles outlined by Deegan and Rothwell (2010) who state that content must be responsive to the smaller screens found in smartphones. This second approach seems to be the most logical as it follows the principles of HCI and web design (Bohyun, 2013). The latest version of the hypertext markup language used in web design (HTML5) allows content to be responsive to screen size and natively allows the embedding of multimedia. This means that content can be dynamically rearranged to match the screen size. These features relieve the developer from having to adapt materials from previous platforms for mobile use (Snell, 2013). Wang and Shen compiled a wide-ranging report in 2012 that set out to synthesise a plan for m-learning material design based on current research in the field. The focus of their paper was “message design” in this context a “message” is defined as “a pattern of signs (words, pictures, gestures) produced for the purpose of modifying the psychomotor, cognitive or affective behaviour” and design referring to the “deliberate process of analysis and synthesis that begins with an instructional problem and ends with a concrete plan or blueprint for a solution” (Fleming and Levie, 1993, p.10). In other words, it involves considering the content of the message and equally importantly how it is delivered.

Multimedia content has been improved over the years in consideration of the message-design principles of Mayer and Moreno (2005), but Wang and Shen argue that Mayer’s theory was developed before the recent uptake in m-learning and has not been tested in this context. Their paper seeks to provide m-learning developers with some underpinning theory on which to base their pedagogical approach. Having taken into consideration some of the theory already covered in this review, such as dual-coding, formal and informal learning, trends and challenges in computer power, network speed, software development, varying device types and platforms, Wang and Shen condense the theory into principles of message design relating to each of these main areas. There is agreement with the findings of the Sølvsberg and Rismark study (2012), in that videos should be kept short - less than 5 minutes long and content should be chunked and

designed to work across different devices. Surprisingly, (and in concurrence with Chu, 2014) Wang and Shen advocate the continuing development of materials for the outmoded PDA platform.

Despite their concerns regarding the lack of testing of Mayer's (2012) theory of Multimedia Learning in the context of m-learning, Wang and Shen continue to support its use in some aspects of content design. Mayer proposes that materials should be coherent, avoiding extraneous content, should provide cues on how to process the information-given and employ strategies recognised to reduce cognitive load. These include the principles outlined earlier in this review, namely: spatial and temporal contiguity, where words are presented in close proximity to related graphics and at the same time and redundancy, where printed words should not be simultaneous narrated. The authors also make an interesting observation about the use of audio relating to the fact that it can be used as an input as well as an output in m-learning, not just for communication with peers, but also as a speech-to-text data entry method. Wang and Shen state that text intended for small screens should be legible in terms of font choice colour and size and should be presented on a non-distracting background. Chunking may be facilitated by the use of colour to link logically related elements and to focus attention on relevant content that might otherwise have been overlooked by the learner. Colours should also be chosen to be consistent with the instructional message and should relate to the intended audience. (Pett and Wilson, 1996)

In terms of developing materials across different device sizes, it is also important to avoid long tracts of text that require scrolling. In a recent study (2014), Molina, et al., undertook two experiments to ascertain the effectiveness of m-learning devices in the delivery of learning materials. The first study compared desktop computers and smartphones and a follow-up study included the use of tablet devices. The aim of the experiment was to investigate the ability of these different technologies to display learning materials. The method used was very thorough, using pre/post-testing, self-reported cognitive load and eye-tracking to ascertain cognitive activity as well as ascertaining user satisfaction scores via the *Technology Acceptance Model* (TAM), a score commonly used in information technology. The hypotheses posed were that the student performance would be

influenced by the devices used and that the time spent in assimilating the learning contents would also be influenced by the type of device used to access them. The third hypothesis did not relate strictly to CLT but looked at student satisfaction and the perceived usefulness of the devices.

The first experiment used 26 participants; the second experiment used an additional 10 participants using iPads (Apple Inc.). The second study combined the data from both trials. The results agreed with Raptis et al. (2013), in that the task time was significantly increased in the m-learner group using smartphones, and was related to screen size. The PC users spent the least amount of time on the learning task and the tablet users task-completion times lay between the two. The reason for this was again due to the time taken in scrolling or pinch-zooming content on the mobile devices that could be displayed without interaction on the PC. This result agreed with the cognitive load reported by the participants in relation to the demands placed upon them by the device, the iPhone users reporting the highest task load, then the tablet users, the PC users reporting the lowest demand. It was thought that this increase in load was likely to be due to the reasons previously evaluated from the Martin and Ertzberger paper (2013), i.e. that screen-size limitations necessitated the splitting of content across different screens, which compromised the spatial and temporal-contiguity principles that result in the split-attention effect.

Interestingly, cognitive load was also assessed relating to the demands placed upon the learners by the learning materials, but there were no significant differences noted here. In agreement with the Martin and Ertzberger study, learners using the iPad tablet found the mobile device to be more motivating than a PC despite the fact that it did not perform as well in relation to the task outcomes. This motivating/user engagement aspect of m-learning is seen in many other studies (Liu, Li and Carlsson, 2010;

Manuguerra, Petocz, 2011; Hwang, Yang and Wang, 2013; Sung and Mayer, 2013; Chan, et al., 2014) Student engagement is one of the factors that can affect cognition because, as stated by Robert Gagné (1985) the first step in any student instruction is to gain attention. In addition to addressing the requirements of CLT, the learning materials must therefore also foster student engagement, as this has also been found to affect cognition.

2.6.4. Engagement and Adoption

One of the first studies to recognise the potential of the tablet-computer as a tool to foster student engagement (identified on page 39, *User Engagement and Empathy*) was undertaken by Manuguerra and Petocz (2011, p.62). Student engagement lies very much in the affective domain, and the approach taken by the study is a report looking at the qualitative aspects of iPad use after an “intense and continuous” 15-month trial of the device at Macquarie University, Sydney, Australia. Manuguerra reports that in the classroom, the devices had been used to “engage inspire and motivate students” through their ability to present high-level presentations and facilitate communication and note-taking, but it was in the field of distance learning that the advantages of the devices were most appreciated by the students. The university had the facility to record live lectures, including all of the presentations and live annotations made by the lecturer, students wishing to watch these recordings had been previously required to attend the campus to do so, and only a few of the largest lecture theatres were capable of screening the recorded lectures. The iPad changed the way students accessed recorded lectures, because it was capable of screening this content remotely, allowing distance learning students to engage with the lectures and benefit from the whole lecture experience in much the same way as they would on campus. This had a measurable effect on student engagement with the course as a whole, leading to a drop in attrition from 15/36 in the year before the iPads were introduced to just 4/37 after the m-learning option was put into place. The authors recognised that mobile devices cater to the expectations of modern learners who have grown up with this technology and are familiar with concepts such as social networking and mobile technology.

The student-perspective on m-learning has been investigated in an in-depth qualitative study by Gikas and Grant (2013) and presents a slightly more balanced view. The study, conducted at three North American universities, used semi-structured focus-group interviews that were transcribed and appended with information about non-verbal behaviours exhibited by the nine participants during the group interviews. Inductive data analysis was employed, and iterative encoding allowed the authors to identify overarching themes and categories. The emerging data revealed positive and negative aspects relating to how the students engaged with their devices. The aspects that

students found appealing were largely in agreement with the findings of Martin and Ertberger (2013). These aspects included the *always-on* connectivity and accessibility of the devices, allowing access to information quickly. The content accessed included discussion boards, course readings and video. Importantly the participants noted that this was not a one-way process, and they used the devices to upload content and make posts on the course site. Students also valued the ability of the devices to screen content uploaded by their lecturers prior to lessons. This is an example of where m-learning can be used effectively in a classroom setting. Another benefit identified by the participants was the communication possibilities afforded by the devices. Students found that they communicated with each other more through video-conferencing applications such as Skype (Microsoft Inc.), Short Message Service (SMS texts) and social networks such as Twitter. This advantage of m-learning was also identified by Manuguerra and Petocz (2011), and Martin and Ertzberger (2013). Twitter was used to share thoughts and exchange ideas with fellow students in preference to the university discussion boards as it was thought to be more immediate and embedded into everyday life. Advantages relating to situated learning were also reported such as the ability to upload photographs and videos to a web-log (blog) while out on the street, rather than having to remember to email them to a tutor at a later time. Holley and Dobson (2008) also identified the use of SMS as a tool to foster student engagement when on a field trip. Students using this feature of their mobile devices to facilitate discussion about art while in the Tate Modern Art Gallery.

This type of seamless learning experience is fostered by mobile devices and has been identified by authors such as Huang, Huang and Wu (2014) and Huang et al. (2011) as an aid to cognition. Without the support of the device, students would be required to hold a mental representation of the context in working memory and this *representational holding* can cause a cognitive overload when combined with other types of mental information processing. Shadiev et al. (2015) relate this to the previously-mentioned *Distributed Cognition Theory*, a socio-technical system whereby representational holdings can be converted into artefacts (such as photographs created with a mobile device). This reduces cognitive load as the user is essentially off-loading some of this burden to the device. Land and Zimmerman (2015) go further in stating that multimedia learning

offered by iPads and their ability to take photographs supports deeper engagement in learners in the field and promotes engagement with their surroundings rather than the “heads down” engagement with the device.

However, the Gikas and Grant (2013) study also identified some barriers presented to student engagement by m-learning. Not all of the issues related to the devices, anti-technology instructors (lecturers) were identified who were either unwilling or unable to incorporate m-learning technology into their course materials. This is a common theme in m-learning that was also echoed by Deegan and Rothwell (2010), who indicated an interdisciplinary disconnect between pedagogues who are uncomfortable with technology, and technologists who are uncomfortable with pedagogy. Gikas and Grant suggested that in their study, the reason might have been due to a generational difference between tutors and learners.

Pachler, Bachmair and Cook (2010) state that cultural attitudes may affect people’s impressions of mobile technology and cite a 2008 study that found that most teachers see mobile phones as distractions and feel that they have no place in school. This attitude is reflected by a statement made by a student in the Gikas and Grant study (2013, p.23) who said:

“my... instructors don’t even want to see them (mobile phones) in class, I have a film-class and I actually got chewed out a couple of weeks ago because I was looking something up about the film we were watching before it started and my instructor was like “I don’t want to see your phone. Put it away”.

This is despite the fact that the university in question’s policy as a whole was pushing for course-management apps for mobile devices.

The second barrier to engagement identified by Gikas and Grant related to what the authors describe as “device challenges”, however on closer analysis of the paper, many of the issues did not relate to the device at all. Students complained that an online poll was difficult to understand, that a video conferencing tool (Oovoo) did not work satisfactorily and that some mobile apps that did not work as expected. All of these issues are

software-related, not device-related. The one device-related complaint related to screen size, specifically the small keypad on the iPhone which made text input difficult. The final theme identified as a barrier to cognition was the distracting nature of the devices, this related to the “allure of social networking applications that were not being used for class”.

In the light of the topic literature it appears that mobile devices present some unique ways access to learning materials, yet simultaneously some unique barriers to learning. This dichotomy suggests that there may be a need for further research into several areas that relate to the research question for this study. These are evaluated in the following section.

2.7. LITERATURE-BASED JUSTIFICATION FOR THIS STUDY

Mobile devices are a relatively new phenomenon; this raised the question as to whether any research has been conducted to investigate their performance as educational tools in terms of cognitive load. From an HCI perspective, one could hypothesise that there will be an increase in cognitive load compared to a non-computer-based learning activity, but one could also hypothesise that the advantages afforded by the devices may off-set this potential barrier to learning. This is not just an important question from an educationalist-perspective, but also from a software-development perspective. If the high production costs and development-time (Chapman, 2006) related to mobile application design do not result in a beneficial change in learning outcomes it may require a re-think as to how these devices are best used.

With these factors in mind, a thorough search of the social science and computer science databases was undertaken. This search revealed no published experimental studies comparing cognitive load in learners using tablet devices, with that found in learners using traditional non-digital learning materials. There were, however, a number of published papers that looked at m-learning from a cognitive-load perspective, and a subset of these considered equipment design. The limitations of these papers that may implicitly suggest the need for further research are listed below

The few device-comparison studies in evidence were mostly from a screen-size perspective (Findlater and McGrenere, 2008; Kim and Kim, 2012; Raptis, et al., 2013) and two of these studies used simulations rather than mobile devices which compromised ecological validity, and in one case technical equivalence. One study looked at cognitive load relating to touch screens (Ando and Ueno, 2010) but this study only considered data input methods (specifically, the ability to write the Japanese character-based alphabet on a touch-screen). At least one paper identified device issues that did not relate to the device, but rather the software (Gikas and Grant, 2013). Several studies compared touch-screen devices with desktop computers from various viewpoints, but not strictly using CLT (Findlater and McGrenere, 2008; Molina et al., 2014). Lack of data relating to m-learning hardware was also identified in 2013 by Raptis et al., who note that previous research

tends to focus on user experience rather than the unique inherent physical attributes of the device. In summary, there is little published evidence to suggest that the mobile-device hardware has been assessed from a CLT viewpoint.

2.7.1. Explicit Recommendations for Further Research Made in The Literature

In addition to the implicitly suggested need for further research from the limitations outlined above, some of the papers reviewed made explicit recommendations.

Hollender et al. (2010), and Schmidt-Weigand and Scheiter (2011) proposed that the increasing complexity of virtual learning environments is likely to impact adversely on the cognitive load of the learner and Hollender specifically identifies the need for further research in this area.

Terras and Ramsey (2012) identified five challenges that are specific to m-learning four of which relate to cognition, namely; *the context-dependent nature of memory, the finite nature of human cognitive resources, distributed cognition and situated learning and metacognition being essential for m-learning*. The authors cite CLT as playing a key role in informing the design of m-learning materials. This is due to the fact that extraneous cognitive load may be heavy in m-learning environments, and developers must take this into account. Mayer's Cognitive Theory of Multimedia Learning is identified as a resource that should be used as a guide in creating learning materials that "provide the richness of face to face learning while delivering it over a lean and mobile medium" (Mayer, 2009 p.825). Raptis et al. (2013) identified the need to conduct further research on tablet devices from a screen size perspective. Alsherhri, Freeman and Freeman, (2013) compared three mobile phones regarding the cognitive load imposed by their operating systems and suggested the need for further research into CLT associated with mobile devices as determined by the environment in which they are used. Similarly, Sung and Mayer (2013) Identified the need to look at m-learning in different environments.

Many of the studies included in this review recruited students below the age of 30 years

to participate in the research (Kim and Kim, 2012; Martin and Ertzberger, 2013; Molina, et al., 2014; Raptis et al., 2013). While this offers convenience to the researcher, it may not present a representative sample of distance learners, in a recent demographical survey, 61% of online learners were over the age of 30 years (Aslanian and Clinefelter, 2012). This suggests the need to conduct research on a wider age-group of m-learners. This need was echoed by Liu, Li and Carlsson (2010) who stated that it would be helpful if further research were conducted to investigate m-learning adoption by users of different age groups.

In Summary, this review has shown that there is a body of research to indicate that Mobile devices are uniquely equipped to reduce cognitive load in the learner if the learning materials are constructed according to the principles of good instructional design as informed by CLT. However, in the field of HCI, there has also been much written about the potential for an increase in cognitive load seen when users are required to interact with a computer. This is why CLT and HCI are both are founded on the same theories of cognition (Gagné, 1985).

Hollender, et al. (2010) identified the fact that extraneous cognitive load can be divided into two broad types when undertaking computer-based learning activity. There is the ECL generated by the instructional design of the learning materials, but there is also additional extraneous cognitive load imposed by the interaction between the learner and the computer.

Classroom Learning Activity

working memory capacity optimal germane resources



M-learning Activity

working memory capacity

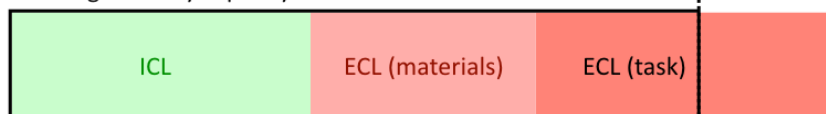


Figure 2-5: (after Hollender et al., 2010; van Merriënboer and Sweller 2010) Proposed ICL and ECL in m-learning

Figure 2-5 shows that ICL and ECL are additive, ECL can be increased by human-computer interaction (task load), If ICL and ECL exceed the capacity of the working memory, germane resources cannot be employed in the formation of schemata. This diagram compares the hypothetical effect of an optimised non-computer-based learning activity on germane resources available to the learner, compared to a similar activity undertaken on a computer.

The Hollender study (2010) looked at human-computer interfaces but was published before the wide adoption of m-learning devices such as tablet computers. This study featured an extensive systematic review of existing published literature and, as such, gives a comprehensive overview of the research at the time. With the exception of screen size, all of the studies analysed the software design rather than equipment design but there were some findings relevant to m-learning, particularly relating to multi-modal learning. This is where a device is equipped to permit learning through multiple sensory channels. In the case of traditional computing, this is usually confined to sound and vision. Mobile devices can extend multimodality to the sense of touch, as they are equipped with a touch sensitive display not commonly found on desktop computers. However, Conway and Christiansen (2005) indicated that when looking at the effect of sensory modalities on learning the sense of touch has been virtually ignored – this still appears to be the case. Furthermore, mobile devices afford situated learning. This aspect of m-learning is known to increase student engagement (Huizenga et al., 2009), encourage informal learning (Chen and Huang, 2012) and provide authenticity to learning that may not be achievable in the classroom (Klopfer, Squire and Jenkin, 2008). All of these attributes can have an effect on cognitive engagement either positive or negative. Taking all of these studies into account, it is proposed that there is a need to investigate cognitive load from a purely device-related perspective. By controlling variables such as software design, material design, and situational context, it will be possible to assess whether the use of the touch-screen device itself has any effect on cognitive load. The proposed study, is, therefore, an experimental two-armed trial to investigate the task

load placed on the learner when undertaking an educational activity on a touch screen device. In consideration of a suitable control against which to test the device, there have been a number of published studies designed to compare mobile devices with personal computers or laptops, but this may no longer be a relevant comparison. Mobile devices were a new phenomenon in the early 2000s, and it might have seemed logical to compare them with the available contemporary technology to determine whether they could be used instead of, or as an adjunct to personal computers. It could, however, be considered that there has been a paradigm-shift away from computers that makes such a choice irrelevant. We are living in the so-called *Post-PC* era, declared by Steve Jobs at the launch of the iPad and reiterated by his successor Tim Cooke at the launch of the iPad Pro (Gilbert, 2015). This is an era in which smartphone adoption outstrips PC use by more than 10 to 1 (Heggestuen, 2013; Ericsson, 2015) and in which half of UK households now possess a tablet PC (OFCOM, 2015). In a recent market-share forecast by the International Data Corporation (IDC) it is predicted that by 2017, smartphones and tablets will take 87% of the worldwide market share for connected devices, leaving PCs with just 5% and laptops with 8% of the market share (Shirer, 2015). Personal computers and laptops simply do not have the ubiquity that is desirable for m-learning. Furthermore, these devices are not equipped for m-learning environments as they are typically not connected to mobile networks. Very importantly, they lack the one essential factor needed for *any-time, any-place* learning – they are not hand-held. At the time of their conception, tablet computers were not intended to replicate or replace computers. Kay (1972) had a very specific concept for the device that was intended to go into production at Xerox Computers, tellingly, the name he gave it was the *Dynabook*. It is therefore suggested that this study should go back to first principles, and look at how learning on a mobile touch-screen device compares against the most appropriate comparator for *any-time, any-place* learning. That medium is not the computer, but the book.

2.8. SUMMARY OF CHAPTER 02

Chapter 02 critically evaluated the current topic literature relating to cognitive load in m-learning. In section 2.1 the FRAME model (Kooze and Ally, 2006) was employed as this offers a structure on which to consider the development of learning materials and hardware for m-learning. A scoping review of the literature revealed that there appeared to be an opportunity to investigate device usability and the learner, an area defined by the authors as contributing to cognitive load in the user. A conceptual map of the pedagogical aspects of m-learning was created to help identify key themes that emerged from the scoping review and this revealed that there appeared to be a gap in terms of how a mobile device might affect the user from an HCI and cognitive load perspective. Section 2.2 looked at how CLT (Sweller, 1989) is used in the field of instructional design, for example when creating e-learning tools. The different types of cognitive load were defined and strategies were evaluated for reducing extraneous cognitive load in the design of learning materials. This theme was developed in section 2.5 where Mayer's (2009) Cognitive Theory of Multimedia was examined. Reducing cognitive load in multimedia presentations is relevant to m-learning as the functionality of mobile devices is congruent with the use of multimedia-rich presentations and apps. It also proved useful in the construction of the learning task in this study, as software related cognitive load needed to be controlled for in the experimental group. In section 2.6 there was a critical evaluation of the literature concerning CLT in m-learning. The four main themes that emerged were device usability, situated learning, material design and learner engagement. Papers relating to these topics were critically evaluated and it was discovered that most previous device comparisons had been made between mobile devices and computers. Some of the papers focussed on software rather than the devices, and at least two of the studies that purported to investigate screen size, used simulations rather than authentic devices (Findlater and McGrenere, 2008; Kim and Kim, 2012). Many of the papers reviewed used convenience sampling and groups whose mean age was not representative of distance learners. (Kim and Kim, 2012; Martin and

Ertzberger, 2013; Raptis et al., 2013; Molina, et al., 2014). Some of the studies used small sample sizes (Raptis et al., (2013); Molina, et al., 2014; Chu, 2014).

Explicit recommendations for further research were made by five of the reviewed papers. These recommendations included the need to investigate cognitive load in mobile-learners, the need to include tablets in studies on screen-size and the need to look at how environment may affect cognitive load in mobile-learners. One author suggested that the sensory modality of touch had been largely ignored in the topic literature. No comparisons appeared to have been made between mobile devices and traditional learning materials such as books.

The conclusion of the review was that there is need for an experimental study to investigate cognitive load in mobile-learners. That a large sample size should be used to achieve statistical power, and that the groups should ideally be representative of distance learners in terms of age and gender balance. Authentic devices should be used rather than simulations and the comparison should be made between mobile devices and books rather than computers or laptops because these devices are not suited to m-learning and, unlike books, are not valid comparators for ubiquitous mobile devices used outside the classroom. Finally, comparisons should be made between task load and device screen resolution as this has seemingly not yet been investigated in relation to m-learning.

3. CHAPTER 03: METHODOLOGY

3.1. INTRODUCTION

The literature review chapter (page 50) provided evidence to support the use of m-learning but also highlighted areas where m-learning may be limited in its ability to deliver learning materials. Many of these concerns relate to the possible shortcomings of computer-based systems, and their known impact on the cognitive load of users – including learners.

Touch-screen devices are still a new, evolving phenomenon, and there have been a number of studies conducted to compare mobile devices with personal computers. However, by design, personal computers and laptops cannot be considered m-learning devices as they are too large to be used *any time, any place* (Martin and Ertzberger, 2013) and do not typically have mobile network connectivity.

This chapter justifies the method chosen to answer the research questions derived from the literature review. It explains the research design, sampling of participants, instrumentation design, the procedure employed in data collection, and a rationale for the data analysis method used.

3.2. DESIGN OVERVIEW

This study employed a cross-sectional, experimental, two-armed controlled trial designed to identify, measure and compare differences in levels of self-reported task load between two parallel, balanced groups of learners during a learning activity. It is recognised that paired tests typically offer more power than independent samples, however this would have necessitated the use of two different learning tasks and two different pre/post tests on every participant. This would have doubled data collection time to approximately 2 hours for each participant and introduced another possible confounding variable as the two interventions would require learning about two different anatomical areas. If the same area were used for both activities, the participants would start the second learning activity with more baseline knowledge because of the prior learning activity. This is the rationale for using independent samples.

The control group were required to undertake a non-interactive learning activity where the information was provided visually/textually, but without the requirement to interact. The learning materials were intended to replicate textbook learning and consisted of a single, labelled, colour photograph showing the structures of the human skull-base. The labels indicated the bony foramina (windows) that serve as conduits for structures (typically for blood vessels and nerves) that connect the brain to the other body systems. In addition to identifying the names of the foramina, the labels also listed the structures passing through each foramen. These are the various cranial nerves, arteries and veins.

The experimental group were required to undertake an interactive multimedia learning activity on a mobile touchscreen device. The content included the same photograph and textual labels but included audio descriptions with a requirement for the learner to interact with the touch-screen. By tapping the structures shown on the screen the learner was presented with a spoken audio identification of the foramina in question, and then the textual information (identical in nature to the text in the control group photograph) was displayed next to the foramina. The aim was to present the same learning materials to both groups, where the only independent variable was the nature of the presentation method.

In each group, the participants were asked to memorise as many of the structures as possible in 10 minutes. Both groups of participants were tested on their pre-existing knowledge of this anatomical area and post-tested after the learning activity to assess what had been learnt (memorised). Pre/post-testing serves as an indirect measure of cognitive load and is explained on page 111. Both groups also completed a subjective rating scale assessment (NASA Task Load Index (TLX)) directly after the learning activity, to measure their levels of self-reported cognitive load and the subscales of mental, physical and temporal demand, performance, effort and frustration. The variables, choice of test used and the rationale for using the measurement tools are explained in the following section of the chapter.

The data were analysed using statistical modelling software to identify any statistically significant relationships between the learning method, the learning outcomes and the cognitive load placed upon the learners by the learning task.

3.3. METHODS CONSIDERED FOR MEASUREMENT OF COGNITIVE LOAD

There are various methods identified in the topic literature that can be used to measure cognitive load; the following section evaluates the suitability of these for use in this study.

Cognitive load relates to internal information processing by the human brain during a particular activity, and is (despite neuroscientific advances), still essentially a theoretical construction. There are, however, various, well-established methods described in the literature that can be used to quantify cognitive load (Pass, van Merriënboer and Adam, 1994). It was, therefore, essential to evaluate these available techniques to determine their suitability for use in the context of the proposed research. To reduce measurement bias, it was considered necessary to use the same tool for both study groups and to prevent the introduction of confounding variables it was essential to ensure that the measurement process did not intrude upon the learning task. For example, a measurement task that requires a participant to use finger tapping (Moskovitch, 1992; 1994; Kane and Engle, 2000), would not be compatible with a learning activity that also requires finger-tapping (such as interacting with a touch-screen). Additionally, a measurement tool that involves anything other than simple user interaction could be self-defeating in having a detrimental impact on the cognitive resources available to the user in undertaking the primary learning task (Brünken, Plass and Leutner, 2003).

The current methods cited in the topic literature for measuring or assessing cognitive load tend to align with two dimensions, namely *objectivity*, and *causal relationship* (Brünken and Plass, 2003).

Objective methods rely on the measurement of physiological responses such as brain activity and task performance (Manoach, et al., 1997; Callicott, et al., 1999; Cowan, 2001; Brünken, Plass and Leutner, 2003; Rubio, et al., 2004; Antonenko, et al., 2010; Durantin, et al., 2014).

Subjective methods are dependent on self-reported levels of task load such as mental effort, frustration and stress (Hart and Staveland, 1988; Kaluga, Chandler and Sweller, 1999; Brünken, Plass and Leutner, 2003; Rubio, et al., 2004).

The causal relationship dimension relates to whether the data collection tool measures cognitive load directly, or indirectly. An example of a direct measurement tool is Functional MRI (fMRI) which allows direct visualisation of brain activation during a task. An example of indirect measurement is the use of pre and post-testing participant knowledge following the learning task. The assumption is that poor performance due to increased extraneous cognitive load will result in a reduced pre/post-test score difference.

The following data collection tools were evaluated for suitability in the context of this research. In addition to objectivity and causal relationships, the factors under consideration included the safety of the participants, the availability and cost of the equipment and the practicality of performing the test on a sufficient number of volunteers required to achieve statistical power in the time available for the study.

3.3.1. Direct, Objective Measurement

The following techniques measure cognitive load directly rather than relying on a second-hand marker and avoid bias that may be caused by (for example) the subjectivity of participant self-reporting.

Functional Magnetic Resonance Imaging (fMRI)

Magnetic resonance imaging (MRI) is well known for its value in the detection of structural lesions, but over the last 20 years, technological advancements in the hardware and software have enabled MRI to be successfully employed in the study of physiological processes such as brain function (Faro and Mohamed, 2010). fMRI can combine the morphological data relating to the structure of the brain, with functional data that relate to brain activity. The principle relies on a phenomenon known as nuclear magnetic resonance whereby naturally occurring hydrogen atoms in the human body can be made to absorb and then release energy - generating a detectable signal in the process (Westbrook, Kaut Roth and Talbot, 2011). fMRI relies on the fact that the brain cortex

metabolises more oxygen when activated than when at rest and allows the researcher to view activation patterns that correspond to cognitive activity. When an area of the brain is in use, the cells burn glucose as part of the metabolic process. This process requires oxygen, as supplied in the haemoglobin of red blood cells. This oxygen is replaced by an influx of fresh oxyhaemoglobin, which provides a subtle intrinsic contrast mechanism that causes a statistically measurable increase in signal from a tissue volume when compared to the same area at rest (Faro and Mohamed, 2010). It has been recently hypothesised that the different regions of the brain are functionally specialised, even at quite a fine level (Rodríguez–Moreno and Hirsh, 2006) and research has shown that functional mapping is possible for various regions of the brain including those involved in cognition. Early fMRI studies demonstrated that the dorsal anterior cingulate cortex and the dorsolateral prefrontal cortex are involved in cognition and play a fundamental role in the mechanism of working memory (Pochon, et al., 2001). More recent research reviewed by Whelan in 2007 has shown that fMRI may also be sensitive enough to be used to identify the parts of the brain that are affected by intrinsic, extraneous and germane cognitive load.

Although the technique is expensive and can therefore only be used on a small sample, it is thought that some effects shown in a single subject are likely to be generalizable, although there will be some effects that are specific to the individual only. Results that are reproduced in more than one test subject are more likely to be generalizable and current research has suggested that a minimum number of participants should be between 10 – 20 to realise a p -value of $\leq .05$ (Faro and Mohamed, 2010). Participant recruitment can be difficult, as suitable volunteers must satisfy certain rigid conditions. For ethical reasons, it is desirable that they have had a previous MRI study of the brain to rule out the possibility of diagnosing incidental pathology during the research scanning. The fact that MRI provides a direct, objective measurement of cognitive load made it a promising consideration for this study, however, there are also many factors that make its use unfeasible. fMRI ideally requires a magnetic field strength of 3T (teslas) or greater; this is approximately 60,000 times more powerful than the earth's magnetic field. As a result, the procedure can pose potentially serious safety issues to research participants. These include contraindication in early pregnancy, damage to implanted devices such as pacemakers and cochlear implants, torque effects on ferromagnetic foreign bodies or

implanted surgical clips and other considerations such as projectile hazards from ferromagnetic items. Powerful magnetic attraction would rule out the use of m-learning devices in the research as these typically contain ferromagnetic components and could become a projectile hazard (Shellock, 2007). Furthermore, MRI examinations deploy pulses of low energy electromagnetic radiation which may under rare circumstances cause burns and damage to implanted devices or the mobile devices being used in the activity. Smartphones and tablets also transmit radio-frequency pulses at frequencies that cause artefactual appearances on MRI images (Westbrook, Kaut-Roth and Talbot, 2011). From a research methodology viewpoint, a potentially confounding factor in using fMRI to measure CL is that the MRI environment is, in itself, very distracting. The scanner has high acoustic noise levels - in excess of 100 decibels, and the volunteer would be required to be confined, lying flat in a cylindrical scanner-bore, with a large detector coil positioned around the head (Westbrook, Kaut Roth and Talbot, 2011). These conditions make interaction with the mobile device difficult, and they present a very artificial environment, quite unlike the usual study environment for a student using a touch-screen device. These factors may make the results less generalizable and also affect the ecological validity of the study. Finally, because MRI poses a risk of discovering incidental pathology in the participant, all of the images taken would require to be reported by a radiologist. Image reporting would have added a great deal of further expense. For these reasons, it was considered impractical and ethically questionable to use fMRI for this study.

[Electroencephalography \(EEG\)](#)

In what is otherwise a diminishing clinical field, the evaluation of neurocognition by EEG has been described by Schomer and Lopes Da Silva (2011) as the most fascinating aspect of modern practice in this field, largely because of recent technological developments that allow EEG to compete with fMRI and other more complex techniques such as single photon emission tomography (SPECT). Antonenko, et al. (2010) concur with this view stating that being able to spot subtle fluctuations in CL with a high degree of temporal resolution (i.e. the ability to witness effects in real-time) can help to explain the effects of interventions that could not be assessed using self-evaluation post-testing. However, EEG

measurement can be a time-consuming and inconvenient procedure. When undergoing traditional clinical EEG techniques, the subject is typically required to wear 20 (or more) electrodes, which adhere to the scalp with a conductive gel. Application and removal of the electrodes is a time-consuming process, and requires preparation by the participant (washing the hair before and after the test). Multiple-electrode EEG is therefore impractical for a large sample of participants. In 2010, Haapalainen, Kim and Dey discovered that it was possible to make EEG recordings using a single dry electrode worn as a headset. The sensor used on such devices makes use of a new technology that uses noise cancellation, digital filtering and amplification to provide high-quality, research-standard EEG without the need for multiple wet-conductive electrodes. Single-electrode EEG is a quicker, less obtrusive technique, and as the electrode is connected via the Bluetooth wireless protocol, there are no wired connections. The above considerations combined with the relatively low cost of the device and minimal safety hazards made EEG an appealing data collection method for measuring cognitive load, where obtrusive external influences may affect the validity of the study. A dry-electrode EEG set was purchased for testing, but it was discovered that sensitivity of the single electrode did not appear to be high enough for the requirements of the test. This technique requires individual testing of participants and it would have been impractical to test the number required to achieve statistical power in the time available. For these reasons EEG was rejected as a data collection method for this study.

Dual-Task Technique

This method requires the participant to perform a secondary task (such as regular finger tapping) simultaneously with the learning task. The performance in the secondary task being a measure of how much load is being placed on the individual by the primary task. (typically an increase in reaction time is seen in the secondary task with increasing primary task load) (Brünken, Plaas and Leutner, 2003).

Drawbacks to this method include the necessity to be able to measure reaction time accurately and the fact that finger tapping (for example) is obtrusive to the primary task. Brünken, Plaas and Leutner identify this as a limitation stating that even a simple secondary task may affect the learning outcomes of the primary task. This technique also requires individual testing of participants, and it would have been impractical to test the

number of participants needed to achieve statistical power in the time available for the study. For these reasons, this method was discounted.

[Near Infra-Red Spectroscopy](#)

This technique uses optodes (“optical electrodes”) to record the haemodynamic activity in the pre-frontal cortex of the brain relating to oxygenation levels. Near infra-red spectroscopy is a safe, non-invasive technique that shows cognitive load in real time and can be used in any environment. It also shows the areas of the brain cortex that are in use (Durantin, et al., 2014). It was not considered suitable for this study, as the equipment is highly specialised, expensive to hire, and requires a high level of user training. This technique also requires individual testing of participants and it would have been impractical to test the number of participants required to achieve statistical power in the time available for this study.

3.3.2. Indirect, Objective Measurement

These are techniques that rely on second-hand indicators of cognitive load and that provide objective data.

[Electrodermal Activity and Heat-Flux Measurement](#)

These non-invasive methods use electrodes, typically on the fingers, to measure changes in skin conductivity (Setz, et al., 2010) or heat flux (Haapalainen, Kim and Dey, 2010). Theoretically, these responses should increase with cognitive load, but there have been studies where the opposite effect has been noted for skin conductivity (Ikehara and Crosby, 2005) which casts the test validity into doubt. Other causes of error include the slow inherent response rate of the tool and artefactual error due to participant motion. This technique also requires individual testing of participants and it would have been impractical to test the number of participants required to achieve statistical power in the time available for the study. Furthermore, finger electrodes would be too obtrusive in a task that requires the manipulation of a mobile device by the participant. These tools were rejected for these reasons.

Eye Tracking and Pupil Size

This method uses a camera to track the gaze of the participant during the test and measure pupil dilation. The test relies on studying eye fixation on (for example) the components of the graphic user interface of a computer during the task. An increase in mental load being reflected in the increasing number of eye fixations (Nakano, 1971). This technique was also rejected, as it requires individual testing of participants and it would have been impractical to test the number of participants required to achieve statistical power in the time available for this study.

Heart Rate Variability

Heart rate variability can be used as a measure of cognitive load (Haapalainen, Kim and Dey, 2010). This test requires the participant to be fitted with electrocardiogram electrodes that measure high frequency (HF) and low frequency (LF) components of the electrical activity generated by the heart during a task. Changes in cognitive load affect the ratio between these components during a task. The reasons that this test was excluded from use in this study was because of the time constraints of having to assess a large number of participants individually, and the ethical risk of finding incidental cardiac pathology.

Performance-Outcome Measures

Before the introduction of the NASA Task Load Index (described later), the most common method to investigate cognitive load was the use of pre and post-testing to assess learning outcomes. Brünken, Plass and Leutner (2003) state that this design is often applied in the context of multimedia learning, typically when comparing two variants of multimedia instruction that use the same material. Because the information content is equal for both arms of the study, intrinsic cognitive load can be presumed to be the same for both groups of learners. Therefore, it can be presumed that any difference in knowledge acquisition can be attributed to a decrease in extraneous cognitive load during the task. Although there can be confounding factors, such as bias generated by differences in preferred learning methods in the participants (Mayer, 2012), there are

other advantages that make performance-outcome a suitable method for consideration. The ability to pre and post-test a large number of learners simultaneously permits data collection from the requisite number of participants in the time available for the study, and the option to use online testing offers the potential to recruit a large number of participants.

For these reasons pre and post-testing was selected as an objective measurement of cognitive load for the proposed research. (Online testing was not used for the final study).

3.3.3. Direct, Subjective Measurement

The following technique purports to measure cognitive load directly and provide subjective data.

Post-Task Questionnaire

This technique described by Brünken, Plass and Leutner (2003) is very similar to the indirect subjective methods covered in the following section. However, it may only focus on the overall mental load perceived by the user who is required to rate the difficulty of learning materials. Moreover, despite being described as a direct measurement of cognitive load, the data can be affected by factors such as learner competency and task difficulty (Kaluga, Chandler and Sweller, 1999; Brünken, Plass and Leutner, 2003). The test also only considers the broad dimension of mental load. For a more sensitive approach, there are multi-dimensional tools available for use in indirect, subjective measurement of cognitive load. It was decided to reject a simple post-task questionnaire in favour of a multi-dimensional tool.

3.3.4. Indirect, Subjective Measurement

These techniques do not measure cognitive load directly but rely on second-hand user reporting of (subjective) data.

Post-Task Subjective Rating Scale

In this data-collection tool, learners self-report dimensions of task load, typically the amount of mental effort that they feel was invested in the learning task. This technique was first identified for use in the investigation of cognitive load in 1994 by Paas, van Merriënboer and Adam. It is safe, non-invasive and because it is administered after the event, it has the benefit of not being obtrusive to the task being measured. The authors state that cognitive load can be defined as a multi-dimensional construct. The dimensions relating to the task and the learner include measurable concepts, namely mental load, mental effort and performance; mental effort reflecting the cognitive load being expended on the learning task. There can also be subjective dimensions such as user-frustration and fatigue. To measure these dimensions, rating-scales are often incorporated in post-task questionnaires. Their use is based on the assumption that learners are able to self-examine and self-report on their mental processes. Self-reporting seems to present the opportunity for introducing threats to construct validity, such as hypothesis-guessing by the participants dishonesty or inaccuracy in reporting, response bias and evaluation apprehension during the test, however Paas explains that there is research to show that participants are quite capable of placing a numerical value on perceived mental burden. This view is echoed by Haapalainen, Kim and Dey (2010) who state that there are many studies to show that self-reporting of cognitive load is a relatively reliable method. They also state that the most commonly used tool is a post-task questionnaire developed by Hart and Staveland (1988) on behalf of the Human Performance Group at the American National Aeronautics and Space Administration (NASA).

The NASA Task Load Index (TLX) was developed to look at the task load imposed on (for example) pilots who were required to interact with technology under stress. Like Paas et al. (2003), Hart and Staveland theorised that mental workload was not unidimensional

and that six different facets could be measured. Their tool is, therefore, a multi-dimensional assessment that is designed to measure task load under the sub-scales of mental demand, physical demand, temporal demand, performance, frustration and effort. Effort, temporal demand and performance specifically relate to cognitive load. The other dimensions provide a wider set of data including physical interaction. The advantage of this tool is that it can be easily applied, can be administered on paper or online and can be given to multiple participants simultaneously. Hart (2006) reviewed the use of the NASA TLX after 20 years of use in research and noted that some modifications had been made to the design of the tool by various researchers over the years. The two most common adaptations are the elimination of the weighted version of the data, and the analysis of the subscales (dimensions) as separate entities. Both of these strategies have been considered for this study, in addition to the normal weighted version of the scale.

For these reasons, and because of its application to human-computer interaction, the NASA TLX was chosen as the indirect CLT measurement tool for this research project, and may allow triangulation of findings between the direct measurement obtained from pre and post-tests.

3.4. DESIGN RATIONALE

Cognitive load (and net task load) caused by a learning activity can be quantified using self-reported data or by evaluating pre and post-test scores. The quantitative nature of the data supports a post-positivist, quantitative approach to the research design (Salkind, 2008). The intention to generalise findings to a target population requires an experimental approach with the ability to employ inferential statistics from the results of the data analysis. This study employed a cross-sectional, experimental, two-armed controlled trial designed to identify, measure and compare differences in levels of self-reported task load between two parallel, balanced groups of learners during a learning activity. Strictly speaking the design could be described as quasi-experimental in that it features the usual structure of a controlled trial in comparing two groups, but differs from a randomised controlled trial in that consecutive sampling was used.

A cross-sectional study was employed, as the research question relates to the present moment. Pachler (2010) states that there is a real challenge when hardware and software become constantly out-of-date. In research terms, longitudinal studies involving technological devices may, therefore, be complicated by the fact that hardware is replaced/updated regularly. It is recognised that a longitudinal study over a protracted time period may introduce a confounding variable or chronology-bias whereby technological advances in hardware design, such as improvements in processor speed or touch screen sensitivity could hypothetically have a bearing on the degree of cognitive load imposed on the user of the device. For example, an assumption could be made that a modern, more responsive device would result in reduced levels of frustration or physical demand to the user. These are two of the dimensions measured by the NASA TLX tool. For this reason, a data collection period of no longer than 12 months was employed. Two-tailed hypotheses were employed as there was no presumption that the mobile-learners group would experience greater, or lesser levels of cognitive load than the non-interactive learners.

The control group were required to undertake a non-interactive learning activity where the learning materials were provided visually/textually, but without the requirement to interact.

The experimental group were required to undertake an interactive multimedia learning activity on a mobile touchscreen device. The content included the same learning materials but included audio and there was a requirement for the participants to interact with the touch-screen.

3.4.1. Independent Variables

The independent variables chosen for the experiment relate to the mode of delivery of learning materials. These were:

Experimental group (mobile devices)

- Screen size
- Device type (smartphone or tablet)
- Spatial Resolution of device display

Control group

- Labelled photograph

3.4.2. Dependent Variables

The dependent variables in this study relate to the dimensions of the NASA TLX. And a pre/post test score about anatomical knowledge. The self-reported dimensions of the NASA TLX are:

- Mental demand
- Physical demand
- Temporal demand
- Performance
- Effort
- Frustration

These dimensions can be combined to give an overall task load score or interpreted individually. Scores relating to mental demand and performance can be used as markers for cognitive load.

3.5. INSTRUMENTATION USED IN THIS STUDY

3.5.1. The NASA Task Load Index

The NASA TLX was developed over a number of years by research team of Hart and Staveland (1988) for the American National Aeronautical and Space Administration. This tool is a multi-dimensional subjective ratings scale that requires the participant to indicate the perceived load placed upon them from six aspects of a learning activity, these are described below.

DIMENSIONS OF THE NASA TLX

Mental Demand

How mentally demanding was the learning task?

How much mental and perceptual activity was required (e.g. thinking, deciding, calculating, remembering, looking, searching, etc.) Was the task easy or demanding, simple or complex, exacting or forgiving?

Physical Demand

How physically demanding was the learning task?

How much physical activity was required? (e.g. pushing, pulling, turning, controlling, activating, etc.) Was the task easy or demanding, slow or brisk, slack or strenuous, restful or laborious?

Temporal Demand

How much time pressure did you feel due to the rate of pace at which the tasks or task elements occurred? Was the pace slow and leisurely or rapid and frantic? Was the task easy or demanding, slow or brisk, slack or strenuous, restful or laborious?

Performance

How successful were you in accomplishing what you were supposed to do? How successful do you think you were in accomplishing the goals of the task set by the

experimenter (or yourself)? How satisfied were you with your performance in accomplishing these goals?

Effort

How hard did you have to work to accomplish your level of performance?

How hard did you have to work mentally and physically to accomplish your level of performance?

Frustration

How insecure, discouraged, irritated, stressed and annoyed versus secure, gratified, content, relaxed and complacent did you feel during the task?

In their original design for the NASA TLX, Hart and Staveland (1988) recommended that each of these subscales should have a wide range of increments. Their rationale being, that scales having fewer increments demonstrate lower sensitivity to experimental manipulations. The original design used a visual scale with an unmarked continuum that could be partitioned into user definable intervals at the time of data analysis (0-100 in their initial validation of the scale). Many modern versions of the test have a scale partitioned from 0-20 and some researchers, such as Haapalainen, Kim and Dey (2010), have used scales with fewer increments (1-5). Hart and Staveland explain that having discrete partitions allow testing when the participant is required to verbally state levels of task load using whole numbers, and fewer partitions make the data easier to sort. For this study, it was not necessary for the participants to verbalise levels of load, so a graphic scale was used. The scale featured 20 marked partitions, but it was explained to the participants that a crossline could be placed at any point along the scale. The width of the scale (16 cm) provided enough space to assign a score between 0 and 100 offering high sensitivity.

Each dimension was rated using this 100-point subjective rating scale. The score can be used raw or can be weighted. Weightings adjust the raw score by the task load dimensions identified as having had the most influence on the activity under investigation. Weighted scores are obtained by the use of a paired-comparison task whereby the participant is required to choose which dimension was more relevant to the

learning task from 15 sets of pairs in which every dimension is paired in combination with every other dimension. A weighted score can then be calculated by multiplying the raw score for each dimension by the weighted score and then dividing by 15 (as there are 15 choices of pairs).

The weighting technique is useful in corroborating the data obtained from the subjective rating scales, for example in this study one control group participant indicated a score of 92/100 for the physical demands placed upon them by studying a labelled photograph. This seems an unreasonably high score for a passive activity, given that the mean score from the rest of the group was 16/100. However, when the same participant completed the paired-comparison task, the weighting given to physical demand was zero (i.e. they did not choose physical demand from any of the pair-wise choices and therefore considered it to be the least influential of all the dimensions relating to task load). The weighted result for physical effort gives a more realistic impression of the user-experience than the raw data in this instance. This is in agreement with Hart (2006) who states that weighting increases user sensitivity to variables and also increases inter-rater reliability. For these reasons, the weighted scoring method was used in this study.

3.5.2. Rationale for Using NASA TLX in this Study

The rationale for using the NASA TLX in the context of this study is based on the following features:

- Subjective measurement scales have been demonstrated to be sensitive to small differences in cognitive load and are valid, reliable and unobtrusive (Paas, et al., 2003). Subjective measures of cognitive load have also been found to correlate highly with objective measures (Paas and van Merrienboer, 1993; Kaluga, Chandler and Sweller, 1998).
- It is considered to be more reliable than other workload measures, to have the highest factor validity (namely the highest correlation with the factor it was intended to measure) and a high test-retest reliability (Noyes, Garland and Roberts, 2004).
- It is a widely used and well-validated technique in assessing the workload in operators of “human-machine” systems such as mobile technology (Hart, 2006).

- It is now considered a foundation of cognitive load measurement and as such, is often used as a benchmark against which other methods are judged (Hart, 2006),
- It has been identified as the most commonly used tool for the assessment of cognitive load due to its ease of implementation and low intrusiveness to the activity being assessed (Haapalainen, Kim and Dey, 2010).
- It is safe and non-invasive for the participants.
- It provides numerical data that can be used with statistical tools required for hypothesis testing.
- It measures additional task load dimensions in addition to cognitive load.

3.5.3. Pre and Post-Testing

A well-accepted method of measuring cognitive load objectively is to use performance outcome measures (Brünken, Plaas and Leutner, 2003). These can include pre and post-testing of knowledge following a learning activity. For test validity, it is important that any additional causes of extraneous cognitive load are either removed from the learning tasks or are present in equal measure in each of the two learning tasks. If additional causes of extraneous cognitive load are present and in unequal measure between the two learning tasks, the results cannot be assumed to be solely due to the variable being measured (i.e. task load imposed by the mode of content delivery). Inferring cognitive load from performance outcome measures, relies on the fact that the information content of the learning materials is the same for both the control group and the experimental group. If the content is identical for both groups, the intrinsic cognitive load can be assumed to be the same for each group. Any difference in post-test scores may be assumed to correlate with extraneous cognitive load induced by the mode of delivery.

Rationale for Choice of Test Topic

The pre-test topic (structures of the base of skull) was chosen because it would be relevant to the target group, requires rote learning and is sufficiently specialised enough to make it unlikely that the learners would have previous knowledge about this area. To assess normal levels of pre-existing knowledge about the structures of the base of skull, the test was trialled in a pilot study on learners representative of the target group. The

mean score achieved was 1.1 out of a possible 36 marks, confirming that it would be unlikely that participants would have a high level of previous knowledge about this anatomical region. The reason that a lack of previous knowledge must be ascertained is twofold, firstly if the participants already understand the topic, there may be little measurable difference in the pre and post-test scores. Secondly, the learning activity may cause extraneous cognitive load due to the expertise-reversal effect as described in section 2.2.2.

3.5.4. Rationale for Using Pre/Post-Testing in This Study

- Before the introduction of the NASA TLX, pre/post-testing was the most common method of investigating cognitive load (Brünken, Plaas and Leutner, 2003).
- Pre/post-testing will provide objective data which may triangulate with the findings of the subjective NASA TLX tool.
- It is ideally suited to a study that compares two variants of multimedia instruction of the same material because the intrinsic load induced will be equal in both variants. If one group of learners acquire more knowledge than the other, it is likely to be due to a comparatively lower level of extraneous cognitive load experienced during the task (Brünken, Plass and Leutner, 2003).

3.6. RESEARCH HYPOTHESES

The two data collection tools permitted 48 (null) hypotheses to be tested, these were:

3.6.1. Between Groups

There will be no statistically significant difference between the experimental group (mobile-learners) and the control group (non-interactive learners) for the following dependent variables:

H1: NASA TLX Net Task load

H2: Net test score between pre and post-tests of knowledge

Subscales of the NASA TLX

H3: reported mental demand during the learning task

H4: reported physical demand during the learning task

H5: reported temporal demand during the learning task

H6: reported (perceived) performance level during the learning task

H7: reported effort during the learning task

H8: reported frustration during the learning task

3.6.2. Between Smartphone Learners and Control Group

There will be no statistically significant difference between the experimental group (smartphone-learners) and the control group (non-interactive learners) for the following dependent variables:

H9: NASA TLX Net Task load

H10: Net test score between pre and post-tests of knowledge

Subscales of the NASA TLX

H11: mental demand

H12: physical demand
H13: temporal demand
H14: performance
H15: effort
H16: frustration

3.6.3. Between Tablet PC Learners and Control Group

There will be no statistically significant difference between the experimental group (tablet PC learners) and the control group (non-interactive learners) for the following dependent variables:

H17: NASA TLX Net Task load
H18: Net test score between pre and post-tests of knowledge

Subscales of the NASA TLX

H19: mental demand
H20: physical demand
H21: temporal demand
H22: performance
H23: effort
H24: frustration

3.6.4. Between Participants Grouped by Device type used (Smartphone or Tablet)

There will be no statistically significant difference between smartphone-learners and tablet-learners for the following dependent variables:

H25: NASA TLX Net Task load
H26: Net test score between pre and post-tests of knowledge

Subscales of the NASA TLX:

H27: mental demand

H28: physical demand

H29: temporal demand

H30: performance

H31: effort

H32: frustration

3.6.5. Between Participants Grouped by Screen Size of Device Used

There will be no statistically significant difference reported between participants grouped by screen-size of device used for the following dependent variables:

H33: NASA TLX Net Task load

H34: Net test score between pre and post-tests of knowledge

Subscales of the NASA TLX

H35: mental demand

H36: physical demand

H37: temporal demand

H38: (perceived) performance

H39: effort

H40: frustration

3.6.6. Between Participants Grouped by Spatial Resolution of Device Display

There will be no statistically significant difference reported between participants grouped by spatial resolution of device-screen used (retina-distinguishable pixels vs. non-retina-distinguishable pixels) for the following dependent variables:

H41: NASA TLX Net Task load

H42: Net test score between pre and post-tests of knowledge

Subscales of the NASA TLX

H43: mental demand

H44: physical demand

H45: temporal demand

H46: performance

H47: effort

H48: frustration

3.7. SAMPLING

Before recruitment took place it was necessary to ascertain the number of participants required for statistical power.

3.7.1. Determining Participant Numbers for Statistical Power

Statistical power can be defined as the ability of a test to detect an existing effect, i.e. the power to correctly reject a null hypothesis. Power increases when the effect is large and test sensitivity increases with sample size (Laerd Statistics, 2016). In quantitative studies, it is an ethical requirement to pre-determine the sample size required to achieve a statistically significant result without having to inconvenience more participants than necessary. To determine the sample size required for a two-tailed test, a calculation was performed using G*Power (3.1) software (Faul, et al., 2009). The following variables were used:

The significance level (the risk of a Type 1 error, whereby the null hypothesis is rejected when it is true) was set at 5% ($p < .05$). The power of the study, (the probability that the null hypothesis will be correctly rejected - avoiding a type 2 error) was set at 80%. These are the usual conventions found in research studies of this type.

t tests - Means: Difference between two independent means (two groups)

Analysis: A priori: Compute required sample size

Input:

Tail(s) = Two

Effect size $d = 0.5$

α err prob = 0.05

Power ($1 - \beta$ err prob) = 0.8

Allocation ratio $N2/N1 = 1$

Output:

Sample size group 1 = 64

Sample size group 2 = 64

Total sample size = 128

Actual power = 0.8014596

Effect size cannot be determined a-priori, however, an estimated or desired effect size can be entered into the calculator. This is commonly set at 0.5 as this provides a moderate effect size. Choosing a lower value could result in a sample size that produces a statistically significant result having a small or trivial effect size, which would be undesirable. The calculation shows that a minimum number of 128 participants (64 participants in each group) are required to achieve the statistical significance, and power required.

3.7.2. Sampling Method

Target Population

The target population was intended to be representative of life-long distance learners over the age of 18 in the field of medical or para-medical healthcare education, for example, doctors, nurses, radiographers, physiotherapists, veterinary surgeons, and researchers.

Sampling Method

Participants were invited from a pool of individuals attending study days on the topic of medical imaging. This allowed the data-collection to be performed with the researcher present and assured the participants of full anonymity as the participant ID was only added to the data collection forms post activity. Consecutive sampling was employed. As a non-probability sampling technique, consecutive sampling may not give results that are generalizable to the universal population. However, from all of the non-probability methods, consecutive sampling is considered to give the best generalisability in terms of representing the target population. For this reason consecutive sampling is often used in medical trials (Kendall, 2003). In consecutive sampling, all applicable participants are included in the study until the required number of participants is reached. In other words, every potential participant has a 100% chance of being enrolled in the study. This avoids the types of selection bias that are often associated with non-probability sampling methods such as convenience sampling (Shuster and Powers, 2005). These might include

only sampling participants from a nearby geographical area, or allowing participants to self-select (volunteer) for a study. Consecutive sampling also reduces the opportunity for intentional or unintentional manipulation of the sample, as the researcher is not required to choose participants (Daniel, 2012).

3.7.3. Participant Demographics

The total number of participants was 130, each group containing 65 participants.

The pool from which the participants were recruited was wide in terms of age and geographical location.

Age

The participants ranged from 20-64 years (mean = 36.07 years).

The mean age of the control group = 36.89 years.

The mean age of the experimental group = 35.26 years.

Geographical Location

Geographically the participants came from the United Kingdom, Ireland, Australasia, the Far East, Middle East, Europe and Scandinavia. All participants were fluent English-speakers.

Gender

The gender ratio overall was 90:40 (69% female: 31% male).

The gender ratio was 46:19 (71% female: 29% male) in the control group.

The gender ratio was 44:21 (68% female: 32% male) in the experimental group.

Further details can be found in section 3.7.3 on page 172.

3.7.4. Inclusion Criteria

The inclusion criteria for the study were:

- (All participants) ability to read and write English
- (Experimental group participants) physical ability to manipulate a touch-screen device.

Participants were required to understand English because attempting to undertake the activities without an understanding of English would have introduced an undesirable confounding variable, namely that reading and writing in a second language is likely to have increased task load for reasons other than the learning activity. One participant from the Scandinavian group commented that their English was not of a sufficient standard for the task, and was withdrawn from the study.

3.7.5. Exclusion Criteria

Participants were excluded from the study according to the following criteria:

- Under 18 years of age
- Known to the researcher, or having a professional relationship with the researcher.

The age limit was applied because of ethical considerations and also because the population that the research was intended to generalise for are distance learners who tend to be post-graduates and demographically-speaking 61% of whom are over the age of 30 years (Dabbagh, 2007; Aslanian and Clinefelter, 2012). It was undesirable to use participants who were known to the researcher for reasons outlined in section 3.7.6.

3.7.6. Ethical Approval

An ethics approval form was submitted and approved (see Ethical Approval, p.363).

Although the study involves the use of human participants, the data collection methods do not pose any significant risk to the participants. The participants invited to take part in

the study were identified as having no personal connection with the researcher to avoid researcher-bias and to eliminate the possibility of any power relationships between the participant and the researcher. The learners were not students, and therefore had no academic connection with the researcher or the university. This strategy minimised the possibility of coercion or a feeling of obligation to participate and also eliminated potential repercussions on their continued learning, assessment or employability. Recruitment of participants was discontinued as soon as the required sample-size was achieved.

There were no monetary penalties imposed on the participants for taking part as they were already attending the location used for data collection (as part of an educational meeting) and were provided with the materials necessary for the activity. No other incentives were offered to take part in the research.

3.7.7. Informed Consent

Informed consent (see Appendix D :) was sought from all participants before the data collection process and participants were allowed to withdraw consent at any time. The participants were informed about the purpose of the research, the duration of the data collection activity and an outline of the data collection procedure was articulated.

According to the principles of non-maleficence (Dickenson, Huxtable and Parker 2010), it was explained that the procedure did not present any risks, dangers or discomfort to the participants. The possible benefits of the research were also explained, including the potential benefit to software and teaching-material design as well as the small potential benefit that was bestowed by the learning task. The nature of the task was chosen because it related directly to the learners' own practice.

There was no requirement for deliberate deception in the data collection process (which usually requires a debriefing session), but there was a concluding session that provided the opportunity to thank the participants for taking part. Participants were reminded that there was still an option to withdraw from the study before they submitted their data collection forms.

3.7.8. Confidentiality

All the data collected were stored safely on a secure password-protected non-networked data storage device and were only be accessed by the researcher and project supervisors. No personally identifiable data were collected, and therefore there were no identifiable data that could have been released to a third party. The only exception to this would have been the emergence of data that resulted in the research participant (or others) being put at risk of harm. Due to the nature of the learning task, this did not occur. Participants were assured that in the event of the research findings being published or made public, only aggregate data would be included and that there would be no data published that could identify them as individuals. This fact was also evident from the data collection forms that did not require any identifying information about the participants other than their age and gender.

All data collection, storage and processing complied with the principles of the Data Protection Act 1998 and the EU Directive 95/46 on Data Protection. No personally identifiable data were stored online or on “cloud” based servers.

3.8. LEARNING MATERIAL AND DATA COLLECTION DESIGN

The initially-proposed research design employed an online data collection tool as it was thought to be a potentially convenient way to capture a high volume of data in a short time period. There were limitations to this method regarding the sampling method, the risk of invalid data collection, bias, and the fact that the learning activity was comparing touch-screen devices with computers rather than authentic non-digital learning materials. These shortcomings became apparent approximately three months into the data collection process permitting a modification in strategy and necessitating a different data collection method to be employed.

The revised data collection method involved conducting the activities in a classroom setting and collecting the test data under controlled conditions. The mobile-learners conducted the learning activity on mobile devices and the control group using a labelled textbook-style photograph.

The reasons for including a description of the discontinued data-collection method in this thesis are twofold:

Firstly, it offers a novel approach to data collection, in that it is a self-contained application that was designed by the researcher to deliver learning materials, assess knowledge and also assess cognitive load in a single HTML5 browser-based package. Secondly, although the method was discontinued in this case, it may be of interest to other researchers in the field to discover that a method that seems to have the potential to recruit and gather a large amount of data from a widely geographically-distributed sample of participants may not perform as intended.

It is suggested that these two factors may offer a small contribution to knowledge from a methodological viewpoint.

A full description of the online data collection tool can be found in Appendix A.

3.8.1. Originally Proposed Data Collection Method (Discontinued)

The original design for the data collection tool was a browser-based application permitting data collection to be facilitated online. The application was authored using *Hype* (Tumult Software). The pre-and post-test sections were created using *Formloom* (Yabdab Software Inc.) and *Rapidweaver* (Realmac Software Inc.). These authorware applications allowed the researcher to create two web-based data collection tools using HTML5, the latest version of the hypertext markup language used in web design. HTML5 allows content to be interactive and natively allows the embedding of multimedia such as audio.

It was thought that this approach would attract a wide and varied sample of the target population, geographically and by age and gender. It was also anticipated that a high number of responses could be achieved in a short time frame and that data analysis could be initiated promptly by the automatic population of online databases by the software. Participants were invited from the researcher's professional networks (*MRI in Practice Group* and *Linked In*) and students from the researcher's own institution, but not under his direct tutelage. Ethical approval was sought and granted to invite students from another university (undertaking medical-imaging courses).

Invitations were sent out to approximately 2,400 individuals, 196 of whom agreed to take part in the online data collection. However, after three months of data collection, only 14 participants had attempted the online tasks, and many of the tasks were not completed fully or were completed incorrectly. This led to concerns over bias, and also completion time for the study as well as the inherent non-generalisability of results gathered from small convenience samples. It was thought that there was a risk of attrition-bias and self-selection bias, namely that only individuals with a particular interest in technologically-mediated learning may have completed the activities. Compeau and Higgins (1995, p.129) first described a phenomenon known as computer self-efficacy (CSE) defined as "an individual judgement of one's capability to use a computer". The concern here was that participants having a low CSE were discouraged from completing the online data collection activity, resulting in data being only collected from participants having a high CSE. This scenario would have introduced a potential cause of bias, as the target population cannot be assumed to have a high level of CSE.

The second concern was completion time for data collection. By logging the number of participants undertaking the online data-collection over the first three months of the trial (and extrapolating the declining rate of participation) it became clear that the time required to complete the data collection would have been longer than three years. This would have made the study impractical due to the time constraints of the doctoral academic process.

Reasons for the poor response rate were not ascertainable. Reminder emails were sent out, but the response rate did not improve. After a consultation with the supervisory team, it was thought that the participants might have thought that they were not sufficiently anonymous, having been assigned a participant ID number to provide during the online tasks. This is a feasible explanation because participants are likely to have felt uncomfortable about their lack of knowledge of human anatomy, an area that was (in many instances) related to their profession. Data collected from the on-line tool was discarded.

3.8.2. Revised Data Collection Method (Used in this Study)

As stated earlier, the response rate for the online activity was poor; it was therefore decided to discontinue this approach to data collection. This was regrettable given the fact that the online tool had taken six weeks of development time. Devising a new tactic for data collection offered the opportunity to minimise some of the possible causes of bias or risks to validity that were evident in the online method (covered in the previous section, p. 133). Furthermore, consecutive sampling was employed instead of convenience sampling which decreases the systematic bias associated with the former. The pool of participants was reasonably representative of the target population (Bowers, House and Owens, 2011).

Location

For the revised data collection model the participants were invited from a series of international study days being organised by the researcher. These were scheduled to take place over a 12-month period, and there were a sufficient number of delegates registered for the courses to achieve the group sizes required by the power calculation ($n=64$). The study days were located in Sydney, Australia; Oslo, Norway and Cheltenham, UK. This recruitment strategy provided a geographically-varied pool of potential participants who were all fluent English-speakers and had an interest in medical education and anatomy. It was thought that having the opportunity to explain the research project to the course delegates would result in a higher participation rate, and this proved to be the case with almost a 100% recruitment rate.

The participants were informed about the research project before attending the study days and were asked to bring along a touch-screen tablet and smartphone if they owned one, and wished to take part.

The data collection took place during the course programme by asking those who were interested to remain behind at the end of the afternoon session or to attend 40 minutes early at the beginning of the morning session.

The participant information sheet was provided to the participants, and the researcher gave a short presentation to explain the nature of the data collection task. Those who did

not wish to participate were given the opportunity to leave at the end of this introductory session if required.

[Randomisation](#)

Participants were then randomised into either the control group or the experimental group using a random number generator in Microsoft Excel (Microsoft Inc.). It was noted early in the process that, frustratingly, many of the participants who were assigned to the experimental group were the participants who had not brought a mobile device to use, or had brought a device but failed to bring headphones. It was thought to be a risk to personal data protection to ask the other members of the group to exchange or share devices due to the potentially sensitive nature of the data and photographs contained on such devices. This issue was addressed by the purchase of six tablet computers, two smartphone devices and 20 pairs of headphones that could be provided to participants at further data collection sessions. This strategy ensured that consecutive sampling could be achieved, as nobody needed to be excluded for not bringing a device, and the required number of participants could be recruited in a 12-month period. It also facilitated random allocation into groups as participants who failed to bring a device could be allocated to the experimental group and participants who were required to use a tablet rather than a smartphone could be provided with such. The process used for data collection was as follows:

[Learning Materials](#)

The learning materials for the session were either a labelled textbook-style photograph of a replica human skull base (control group) or the interactive mobile app (experimental group). As the aim of the experiment was to assess the differences between the mode of delivery in each case, it was necessary to use careful instructional design to reduce any causes of extraneous cognitive load that may be due to the learning materials rather than the delivery method.

The learning task that was used for the study required the participants to memorise the structures found in the human skull-base. The options available to distance learners for learning human anatomy are somewhat limited compared to campus-based students

who typically have access to dissection laboratories, histological specimens and realistic anatomical models. Mobile-learners are typically confined to using learning materials that they can easily carry, and would therefore traditionally rely on textbooks for learning anatomy. This type of learning falls into the cognitive domain of learning taxonomy (Bloom, 1956) and is, therefore, relevant to the research question, and the data collection methods used. The choice of topic also conformed to good ethical practice as it was relevant to the participants' own educational needs in that it offered some benefit to the participants that would not have been evident had they been required to learn a topic that was of no direct usefulness to their clinical practice.



Figure 3-1: Replica human skulls, plastic and resin-cast models

The Human Tissue Act (2004) requires establishments that use real human tissue for teaching purposes to be licenced to use human remains. For this reason, many anatomy classrooms now use replicas. For the learning task, a standard plastic replica skull was purchased to be photographed. On delivery, it was thought that the model lacked the detail required for the learning task, as some of the structures were not replicated well enough to identify. A second model was sourced and purchased. As shown in Figure 3-1, the second replica was a resin-cast model and was indistinguishable from an authentic cadaver specimen.

Labelled photograph

The labelled photograph was a high-resolution image featuring the base of the resin-cast replica human skull. The structures were labelled in accordance with instructional design theory to reduce any sources of extraneous cognitive load (Sweller, 1989; Sweller, van Merriënboer, and Paas, 1998; Mayer, 2009). The text was placed in close association with the corresponding structures to reduce split-attention effects. The names of the nerves and vessels relating to each structure were listed underneath each label. The use of a labelled photograph provided a higher degree of ecological validity than would have been realisable in the discontinued online data collection tool where the photograph would have been presented on a computer monitor and would not have provided an authentic comparator to a textbook diagram. Presenting the diagram in the same format to all learners also avoided the possibility of a confounding variable that may have been intrinsic to the discontinued online version of the data collection. Learners having different screen resolutions and monitor sizes would have experienced the photograph with varying degrees of spatial resolution and size; this is a factor that is known to affect cognition (Raptis et al., 2013; Lin, Wang and Kang, 2015).

Interactive Mobile Application

The mobile application used in the learning activity was developed by the researcher and featured a log-in screen and a second screen showing the learning activity. The login screen featured a “keypad” lock that prevented access to the learning activity until a code had been provided and entered. This strategy allowed the participants to download the

app in advance without being able to open it until they were provided with the code after the pre-test. Pre-downloading was found to be necessary because some of the venues used for the data collection did not provide wireless networking and the mobile network connectivity was often slow. By downloading the app in advance, it was possible to ensure that all of the participants were able to commence the learning activity at the same time without any delays due to download errors. The login screen also featured an audio test button that allowed the participants to set up their headphones in advance to ensure that they could hear the audio component of the app clearly. The second screen showed the learning activity. To reduce confounding variables and to ensure that intrinsic cognitive load was the same for both the control group and experimental group, the learning activity was required to resemble the non-interactive activity as closely as possible, but also feature the interactive functionality that is typically offered by a touch-screen device.

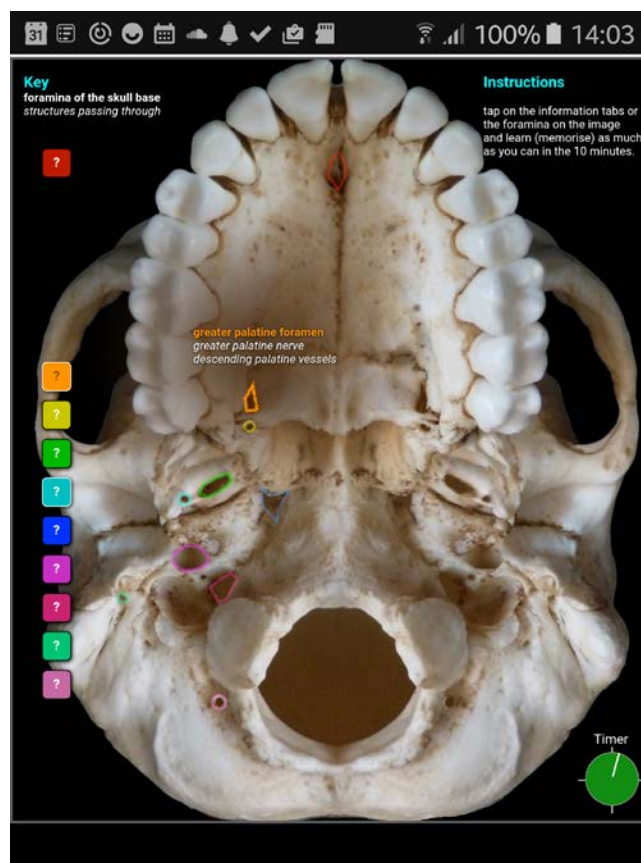


Figure 3-2: Screenshot of the application used in the mobile-device learning-task application (Samsung Galaxy S7 smartphone)

Figure 3-2 (above), shows the mobile app screen after the learner has tapped on the Greater Palatine Foramen (outlined in orange) to reveal the labels. The app plays an audio description of the foramen first and then displays the labelling. The orange outline is emphasised (opacified) by the software to highlight that it has been selected. The task timer is shown at the bottom right of the screen.

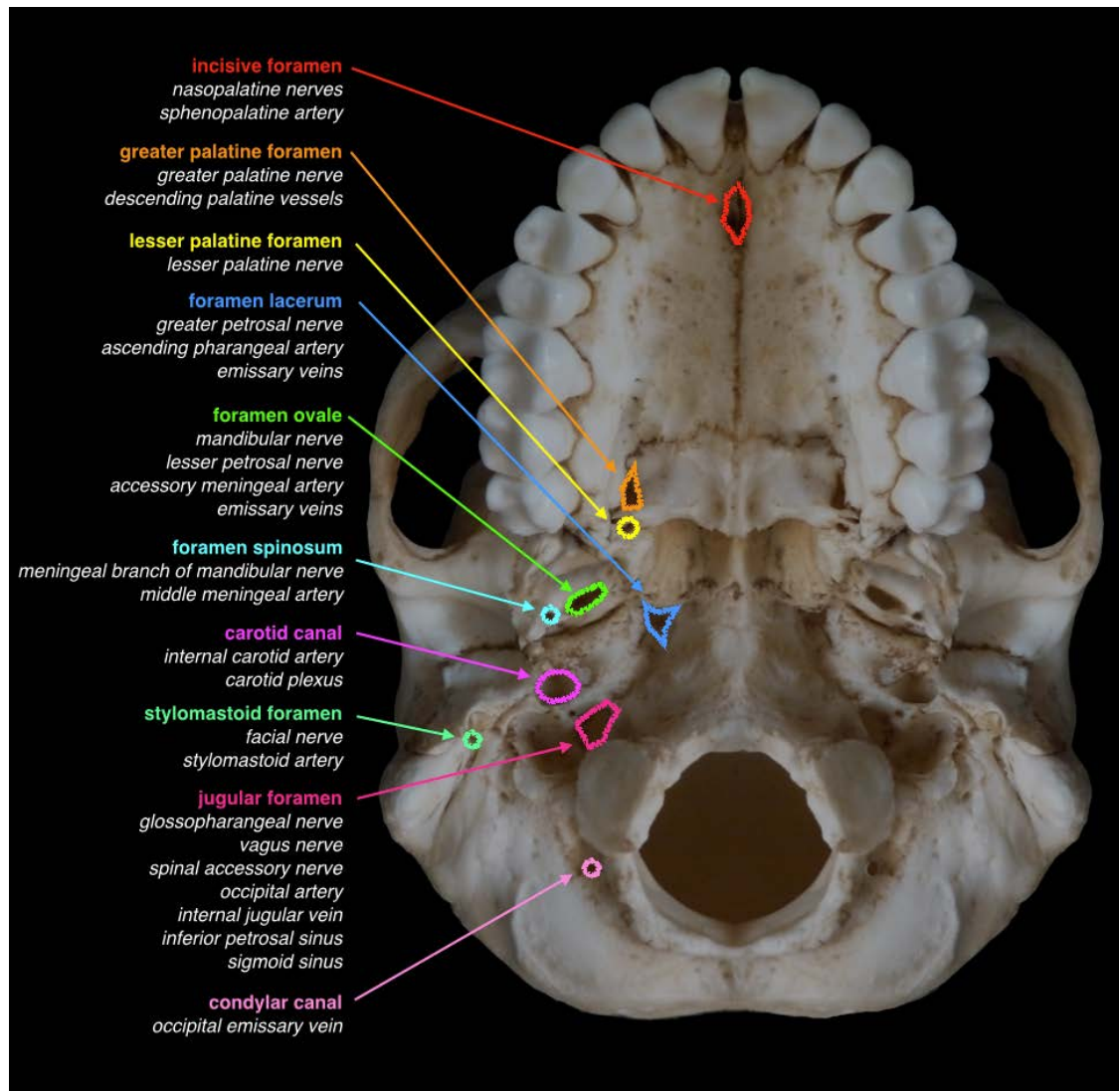


Figure 3-3: Control group learning materials, a labelled photograph

To ensure comparability with the control-group activity, the same high-resolution photograph was used, and the same labels were employed. To activate the labels, the participant was required to tap on the various foramina (bone-windows) on the image. These were outlined in colour and when tapped an audio cue was spoken over the

headphones to identify the name of the foramen in question. Immediately after the audio description, a text label appeared next to the corresponding structure, including the names of the associated nerves and vessels. As for the non-interactive materials, this configuration follows the principles of good instructional design by reducing extraneous cognitive load due to the split attention effect (by placing the label close to the corresponding anatomical structure) and the redundancy effect (by ensuring that the audio and text descriptions were not presented simultaneously (Mayer, 2009; Sweller, 1989).

The rationale behind reducing any potential causes of extraneous cognitive load in the learning activities was to eliminate any variables that did not directly relate to the mode of delivery. Task-completion time, for example, can be an indirect measure of cognitive load (longer task completion times being caused by a decrease in germane resources in working memory (Baddeley, 1974; 2000; Chen, Epps and Chen, 2011). Task-time in m-learning can be affected by other processes such as the necessity for content scrolling on smaller screen sizes (Raptis et al., 2013). Excessive content-scrolling would have presented a confounding variable between the two experimental groups, as the non-interactive learners would only be required to look at a static photograph and not be required to scroll content throughout the task. To reduce the need for scrolling, the interactive activity was designed to be responsive to screen size, namely that the diagram would automatically shrink or expand to fit the available area. Users could use pinching and swiping to zoom into the content as required, the need for excessive scrolling was reduced, but not completely eliminated.

There was a timer on this screen which closed the activity at the end of ten minutes. The rationale for timing the activity was to ensure that both groups had an equal amount of time for the activity. This ensured that any differences in temporal demand could be attributed to the learning task.

3.9. DATA COLLECTION PROCEDURE

3.9.1. Stage 01: Participant Information and Consent

Participants were given a thorough explanation of the research project and the researcher talked them through the participant information form (Appendix D :) and asked if there were any questions. The researcher explained that consent could be withdrawn at any time and that participants could opt not to submit the data collection forms at the end of the collection activity if they did not wish to be included in the study.

3.9.2. Stage 02: App installation

The participants in the experimental group were provided with a link to the online web app described on page 138. The link was provided in advance to allow them to download and open the application in readiness for the learning task.

3.9.3. Stage 03: Pre-Test

The pre-test, NASA TLX and post-test were handed out as paper documents to all of the participants (these can be found in Appendix B :). The pre-test was conducted by screening the skull-base image over the conference projection system and asking the participants to write down the names of as many of the arrowed structures as they could identify in a ten-minute timed session. The rationale for projecting the image was to ensure that none of the participants could access the pre-test content before the task started as this was a timed exercise. The rationale for controlling the time variable was to ensure that all participants in future data collection sessions would have an equal amount of complete the pre and post-tests. During the first data collection session, it was noted that one participant had returned to the pre-test answer-sheet after the learning activity and had attempted to fill in the answers that they had failed to complete during the pre-test. This paper was voided after collection and not used in the data analysis. Following this incident, participants in all future data collection sessions were asked to draw a line through any of the boxes on the pre-test answer sheet that they were unable to complete. This strategy prevented participants from filling the answers retrospectively following the learning activity.

3.9.4. Stage 04: Learning Activity

Control Group

The control group were provided with a labelled photograph of the anatomy to be learnt (see Figure 3-3) and asked to study it for 10 minutes and to learn (memorise) as many of the labelled anatomical structures as possible. Another potential source of data invalidation was spotted at a UK data-collection session wherein it was identified that one participant was writing down the names of the anatomical areas during the learning activity. Having these cues to refer to in the post-test would have invalidated the data, as the answers would not reflect learning. This paper was, therefore, voided and the data not used in the analysis. For all of the following data collection sessions, it was explained to the participants that no writing was permitted during the learning activity.

Experimental Group

The experimental group were asked to open their web apps (see Figure 3-2) and were provided with a pass-code to open the learning activity. They were then given ten minutes to learn the labelled anatomical structures. The code was required to ensure that the experimental group did not open their learning activity before the pre-test had been completed as this would have provided the answers. The experimental group were asked to remain in the same environment as the control group to ensure that there were no confounding variables such as distraction from other sources.

The interactive activity featured the same photograph of the skull-base but allowed the participants to tap on the various foramina. By tapping the screen, auditory and visual information was presented to the learner identifying the name of the foramen in question and the structures passing through. In

Figure 3-2 the screenshot shows the appearance presented when the participant taps on the greater palatine foramen. The coloured outline of the foramen is highlighted (from being semi-opacified to fully opacified) simultaneously with a short, spoken audio description identifying the structure in question. Following the audio description, the software presented a text label next to the foramen (Figure 3-3). The same audio-visual highlighting and labelling could be activated by tapping on each of the foramina shown in

the photograph. The interactivity of the screen also permitted the users to tap the corresponding foramina on the opposite side of the screen (the skull base being symmetrical left to right) and cause the same information to appear. The photograph embedded in the mobile app was presented at a higher resolution than the typical mobile device screen size and was, therefore, zoom-able without loss of detail (by using the default pinching gesture on the screen). There were a set of coloured buttons located down the left of the screen that could be clicked to highlight the foramina more easily than clicking an un-zoomed diagram. These features were intended to aid participants using small screen sizes such as smartphones.

The principles of Cognitive Load Theory (CLT) were applied to the material design to reduce confounding additional extraneous cognitive load imposed on the participants due to the split-attention principle (Nielsen, 1994; Mayer, 2009) or the redundancy principle (Mayer, 2009; Sweller & van Merriënboer, 1998). To avoid cognitive load imposed by the split-attention effect, it was important that the learning tasks should not require the participant to mentally integrate multiple information sources. This spatial contiguity was achieved by ensuring that the diagram and labels for the control-group task were not spatially separate and that the text labels could be viewed simultaneously and in close proximity to the corresponding foramina shown on the photograph (Mayer, 2009). To avoid extraneous cognitive load imposed by the redundancy principle, it was necessary to ensure that images, text and spoken narration/audio were not presented simultaneously (Paas, et al., 2003; Liu et al., 2012). For this reason, the audio cue was presented before the appearance of the text labels in the interactive presentation provided to the experimental group.

The learning activity session was carefully invigilated to ensure that none of the participants wrote down any information that could have been used in the post-test. The control group activity was timed, and the participants were asked to stop the activity after 10 minutes had elapsed. The experimental group activity was automatically timed, and was closed by the app after 10 minutes. It could be hypothesised that the presence of the timer on the screen of the device might have increased extraneous cognitive load if the experimental group had been distracted by it during the activity. This variable was

controlled by presenting the same timer as a large projection on the screen in the data-collection room for the control group to refer to during the task.

3.9.5. Stage 05: NASA TLX Questionnaire

The task load relating to the learning activity was assessed using the NASA Task Load Index. For the NASA TLX, the instructions were screened to the participants over the main conference projection system and the researcher went through the instructions and descriptors before asking the participants to indicate their responses on the paper data collection forms provided. Both sections of the form were used to assess the load placed on the participants by the six dimensions of the tool and also the pair-wise choices to enable weighting of the results. To minimise procedural bias, this activity was not timed, and enough opportunity was given for all participants to complete their responses. One issue arose during the NASA TLX data collection that required papers to be declared void. Some participants did not fully complete both sections of the NASA TLX, usually by failing to indicate one or more of the pair-wise comparisons. To avoid this happening in future data collection sessions, the researcher asked all participants to double-check their sheets at the end of each session to ensure that they had completed all of the sections correctly.

3.9.6. Stage 06: Post Test

The final data collection activity was the post-test. The test was administered under the same conditions as the pre-test. The reason for using a post-test was to allow the pre-test score to be subtracted from the post-test score to give a measurement that could be triangulated against the reported task load to see if there was a correlation to support the assumption that post-testing is an indirect measure of cognitive load. The reason for presenting this test after the NASA TLX rather than directly after the learning activity was to ensure that any knowledge gained during had been committed to long term memory. The activity was invigilated to ensure that participants did not seek to gain an advantage by filling in the pre-test section retrospectively.

After the post-test, there was a text-field on the data collection form to allow the participants to provide an email address (not compulsory) and demographic information relating to their age, gender and the type of device used.

To ensure anonymity participants were asked to fold their sheets and place them in a box when leaving the room. It was explained that if anyone wished to withdraw from the study at this point that they could discard or keep their data-collection sheets rather than submit them to the collection box.

This concluded the data collection activity. The above process was conducted six times over a twelve-month period with groups of approximately 20 participants in each instance.

3.10. METHOD OF DATA ANALYSIS

3.10.1. Descriptive Statistics

Descriptive statistics are presented to summarise the participant demographics and the results of non-statistically significant findings.

3.10.2. Inferential Statistical Tests

Although there are a number of tests described below, there were essentially only three main requirements in analysing the data. Firstly to compare data collected between two independent groups (for example mobile learners vs non mobile learners), secondly to compare data between a number of groups (for example when looking at learners grouped by screen size of device) and thirdly to analyse correlations between variables (for example physical demand vs screen size). For normally-distributed data the statistical tests required are the independent samples t-test, Analysis of Variance (ANOVA) and Pearson's correlation Test respectively. However, some of the data were not normally distributed and non-parametric versions of these tests were also required, namely, Mann Whitney U-test, Kruskal-Wallis test and Spearman's correlation test. Data distribution is covered in more detail on page 159 and the tests chosen are identified below. For tests between two independent groups an assumption is made that the amount of variability in each group is equal. This is required when a t-test or an analysis of variance is required, and when testing hypotheses. To avoid order-effects, independent measures were employed. The participants were randomly allocated into two groups whereby each condition of the independent variable was applied to a different group. The groups were not tested more than once. A number of additional statistical tests were required to test for normality and variance in the data, these are identified in the following section.

3.10.3. Rationale For the Statistical Tests Used

For data that exhibited a normal distribution, parametric tests were employed. For non-normally distributed data, equivalent non-parametric tests were chosen. Some authors such as Lix, Keselman and Keselman (1996) assert that parametric tests such as ANOVA

(analysis of variance) are not particularly sensitive to departures from normality, especially at larger sample sizes (McDonald, 2014) and where the group sizes are equal (Laerd Statistics, 2016). For this reason, for some of the key tests, especially where the histograms appeared to be broadly similar, a parametric test was used to support the findings of the non-parametric alternative.

Parametric Tests

Independent Samples t-test

A *t*-test is used in hypothesis-testing when examining the differences between groups on one or more variables. When there are two groups be included, and the participants are not being tested more than once, a *t*-test for independent samples is indicated. (Hawkins, 2014)

The test has six assumptions:

1. There is one dependent variable that is measured at the continuous level, (i.e. numeric and on a scale that can be infinitely divisible).
2. There is an independent variable that consists of two groups
3. There is independence of observations (i.e. no relationship between the groups, different participants in each group).
4. There should be no extreme outliers in each group regarding the independent variable.
5. The dependent variable should follow a normal distribution
6. The variance of the dependent variable is homogenous (i.e. the same for both groups).

The aim of the *t*-test is to determine whether any differences between the sample means reflect a difference in the population that the samples represent. The findings may be

generalisable and inferences can be made about the population being tested (Donnelly, 2007).

One-Way Analysis of Variance (ANOVA)

This test is used to determine whether there are any statistically significant differences between the means of more than two samples (Hawkins, 2014). The test has what is known as an omnibus test-statistic, in that it can only detect a difference between groups. It cannot specifically identify which groups were different. However, post hoc (follow-up) tests can be performed to make this discrimination. ANOVA makes six main assumptions:

1. There is one independent variable, measured at the continuous level
2. There is one dependent variable that consists of more than two categorical, independent groups.
3. There is independence of observations
4. There should be no extreme outliers in each group in terms of the independent variable.
5. The dependent variable should follow a normal distribution
6. The variance of the dependent variable is homogenous (i.e. the same for all groups).

The ANOVA test is used to determine whether any differences between groups can be attributed to sample error alone rather than variance caused by the independent variable. (Laerd Statistics, 2016)

[Pearson's Correlation Test](#)

Pearson's correlation was used to measure the strength and direction of any association between any two continuous variables. The test makes the following assumptions (Hawkins, 2014):

1. There are two variables measured on a continuous scale
2. The variables are paired
3. There is a linear relationship between variables
4. There are no significant outliers
5. There should be bivariate normality

Pearson's correlation coefficient (r) has a range of values from -1 representing a perfect negative linear relationship between variables, to +1 where there is a perfect positive linear relationship. A value of 0 indicates no relationship.

[Non - Parametric Tests](#)

[Kruskal-Wallis Test](#)

The Kruskal-Wallis test is considered to be a non-parametric equivalent of the One-way ANOVA, used when data do not meet the normality assumption required. This test makes four main assumptions (Hawkins, 2014; McDonald, 2014):

1. There is one dependent variable measured at the continuous or ordinal level
2. There is one independent variable consisting of two independent groups or categories
3. There is independence of observations
4. The data from each group should have a similar distribution

It may be used to test for differences between groups, between conditions, or between change scores. The Kruskal-Wallis test is appropriate for this study when testing for a difference between learners when grouped by the screen size of their devices.

Mann-Whitney U test

The Mann-Whitney U test was formerly known as the Wilcoxon Rank Sum test and provides an alternative to the *t*-test. It is used when the data are not paired, and the data do not follow a normal distribution (Hawkins, 2014). It may be used to test for differences between groups, between conditions, or between change scores. This makes the Mann-Whitney test appropriate for this study where two independent groups were tested under different conditions and a difference in the change between pre and post-test scores was measured between groups. This Mann-Whitney U test makes four main assumptions:

1. There is one dependent variable measured at the continuous or ordinal level
2. There is one independent variable consisting of two independent groups or categories
3. There is independence of observations
4. The data from each group should have a similar distribution (same shape)

The test was chosen as it is thought to offer stronger evidence than other non-parametric tests because it compares the distribution of the samples as well as the medians. Conroy (2012) states that strictly-speaking, the test should not be described as “non-parametric” because it calculates a parameter; namely the Mann-Whitney test statistic. This can be useful in clinical trials, as a measure of effect size when measuring scales that are not interval. Conroy gives examples such as measuring moods and attitudes. The Mann-Whitney U test was, therefore, appropriate for analysing the NASA TLX data collected in this study. The test employs different computations depending upon the sample sizes used and the normality of the distribution of the values in each sample. For smaller

sample sizes (such as found in the smartphone ($n=36$) and tablet-learner ($n=25$) groups) the test calculates a u -statistic. For large sample sizes (such as found in the control group ($n=65$) and experimental group ($n=65$) the value of U approaches a normal distribution and a z -test may be used.

Spearman's Correlation

Spearman's correlation was used to measure the strength and direction of any association between any two variables. The test makes the following assumptions (Hawkins, 2014):

1. the variables can be measured on a continuous, or ordinal scale
2. there are paired observations, namely that a single participant will be affected by the score of two variables, for example, performance and pre/post-test result
3. there is a monotonic relationship between variables, namely that one variable increases (or decreases) with the other.

Spearman's correlation was used in preference to Pearson's correlation where the data were not normally distributed.

Shapiro-Wilk's Test

The Shapiro-Wilk's test (1965) can detect departures from normality due to Skewness and Kurtosis (Razali and Wah, 2011) Skewness describes an asymmetry between the tails of the Gaussian curve, Kurtosis refers to the shape of the peak. Salkind (2008), explains that a peak having a taller, sharper profile than a normal distribution is described as leptokurtic, and a peak having a flatter profile as platykurtic. Kurtosis, in turn, relates to the number of non-typical values seen in data-points from outliers as these will affect the thickness (height) of the tails. A Shapiro-Wilk test was therefore performed on all of the data to determine normality.

3.10.4. Effect Size Calculations

Having determined any statistically significant differences between variables, effect size calculations were conducted to determine the magnitude of these differences. Salkind (2008) states that this is important in gauging whether the differences measured offer a meaningful result.

Effect Sizes for Parametric Tests

ANOVA Effect Size

For ANOVA and t-test data, Cohen's effect size calculation (1988) was used.

$$d = \frac{|M_1 - M_2|}{s_{pooled}}$$

Where M1 is the mean for the experimental group, M2 is the mean for the control group, and s is the pooled standard deviation. The effect sizes stated correspond to Cohen's thresholds for interpreting effect size.

Test	Effect Size	small	medium	large
Standardised mean difference	d	0.20	0.50	0.80
Correlation	r	0.10	0.30	0.50

Table 3-1 Effect size thresholds used for parametric tests (Cohen 1988)

Effect Sizes for Non-Parametric Tests

Mann-Whitney Effect Size

For the Mann-Whitney test, effect sizes (r) were calculated from the Mann-Whitney U statistic using the formula

$$r = \frac{z}{\sqrt{n}}$$

Where z is the Mann-Whitney test statistic, and n is the total number of participants (Grissom and Kim, 2012). z is used instead of U where there are large sample sizes.

Kruskal-Wallis Effect Size

For Kruskal-Wallis tests, effects size η^2 (eta squared) was calculated from the Kruskal-Wallis test statistic (Chi-Square) using the formula

$$\eta^2 = \chi^2 / (N-1)$$

Test	Effect Size	small	medium	large
Standardised mean difference	η^2	0.20	0.50	0.80
Correlation	η^2	0.10	0.30	0.50

Table 3-2 Effect size thresholds for non-parametric tests

Spearman's Correlation Test Effect Size

Effect size was reported as Spearman's Rho (r_s) where a size of 1.0 (or -1.0) represents an identical association between ranked variables and a value of zero represents no association between variables (Laerd Statistics, 2016).

3.10.5. Summary of Decision-Making Process for Statistical Tests Used

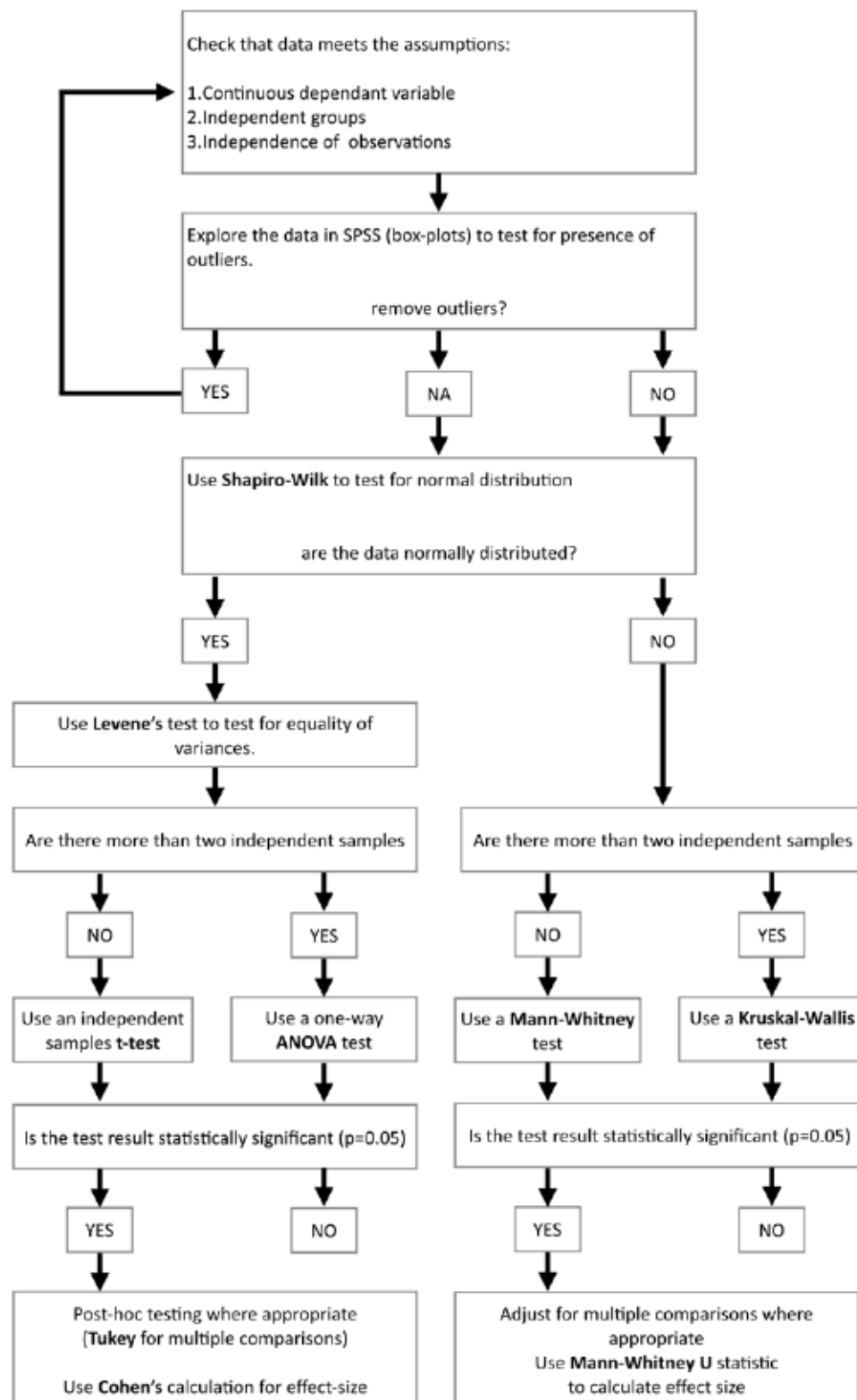


Figure 3-4: Flow chart showing the decision-making process used to determine appropriate statistical tests

3.10.6. Methodological Limitations

Sampling Limitations

The main limitation of the sampling method is that consecutive sampling was used which is non-random and therefore non-probabilistic. The weaknesses of consecutive sampling occur when all of the participants are recruited at a particular time of year (overlooking temporal variation) or if the response rate is low (causing selection bias). Both of these effects were minimised by conducting the data collection over a 12-month period, and by ensuring that all of the invited participants took part in the study. Recruitment was easier to achieve in a face-to-face setting than online, where participant procrastination or lack of motivation might have led to high attrition rates. Over the entire data collection period, only one of the invited participants declined to take part.

Sample Limitations

The required sample size was achieved as there was a wide pool of available participants. The pool featured a large geographic spread and also a wide spread in age. There was an imbalance in terms of gender in the sample, but this seems to broadly reflect the gender imbalance in the target population (UK health-sector workers, 77% female, 23% male (NHS Employers Association, 2015), Diagnostic Radiographers 84% female 16% male (Society of Radiographers, 2009). The mean age of the participants (36 yrs) was closely representative of distance learners (mean age 34 yrs) (Dabbagh, 2007).

Data Collection Tool Limitations

The NASA TLX instrument is a subjective measure, however subjective measurement scales have been demonstrated to be sensitive to small differences in cognitive load (Pass, et al., 2003) and have been found to correlate highly with objective measures (Paas and van Merriënboer, 1993; Kaluga, Chandler and Sweller, 1998) and to have a high test-retest reliability (Noyes, Garland and Roberts, 2004). The results will be triangulated with pre/post-testing which is an objective measure.

Device-Related Limitations

In the mobile-learners group it was intended that tablets and smartphones should be equally represented and that these should be randomly allocated. Most participants owned a smartphone and also brought along a tablet as requested. However, some participants did not own a tablet and had to be provided with one for the learning activity. At the time of designing the data collection activity it was not realised that this could introduce a confounding variable, namely that familiarity with a device, and having a preferred screen-size may affect cognitive load (Raptis, et al., 2013). This is recognised as a possible limitation of this study and is further discussed on page 260. It is not anticipated that device-unfamiliarity had a major bearing on the results of the study, as only a small number of participants used an unfamiliar device. It is unlikely to have affected the validity of the study because m-learning is de-facto undertaken on both smartphones and tablet computers and any inherent properties of these devices, or unfamiliarity with the devices that may affect cognitive load should therefore be recognised as relevant to the outcome.

3.10.7. Delimitations

The study was limited to the target population as described in the sampling section on page 127. The mobile task activity (experimental group) was limited to a classroom environment which may not be typical of the m-learning environment, but this was necessary to control for environmental variables such as user distraction and to prevent invalid data collection due to participant error as described on pages 142 and 143.

General limitations of the study are evaluated in section 6.5. The full research boundaries are identified and justified in section 6.2.

3.11. SUMMARY OF CHAPTER 03

M-learning uses computer-based mobile devices to deliver learning materials. These devices may impose an extraneous cognitive load on the user/learner; a phenomenon that is recognised in the field of HCI and multimedia design. Cognitive load can be numerically quantified by self-reported (subjective) rating scales such as the NASA TLX, or by (objective) pre/post-testing. Numerical data predisposes the use of a quantitative methodological approach to answer the primary research question relating to whether there is any difference in cognitive load experienced by a mobile learner, compared to a learner using a comparatively-sized, ubiquitous learning aid such as a textbook. The requirement to generalise findings to a larger population required an experimental approach.

A cross-sectional, experimental, two-armed controlled trial was designed to identify, measure and compare differences in levels of self-reported task load between two parallel, balanced groups of learners during a learning activity (65 participants in each group). The activity in question required the participants to study a labelled photograph of the base of a human skull and memorise the anatomical features shown. Memorisation represents an example of learning in the cognitive domain and therefore is relevant to the research question and data collection technique. The control group studied a static labelled photograph, the experimental group studied the same photograph on mobile devices and were required to interact with the screen to reveal the labelling for each structure. There was also an audio component, namely the anatomic features were verbally identified when tapped.

The initial data collection tool was an online model, but after a short trial, this method proved to have too many sources of potential bias and issues with validity and poor participant recruitment. The method was therefore improved and adapted to a live data collection environment. Data was collected over a 12-month period in various geographical locations and from a wide demographically-diverse group of participants. The data was collated in a workbook (Excel, Microsoft Inc.) and imported into statistical analysis software; *SPSS* (Statistical Package for the Social Sciences) (International Business Machines Corp.) and *Wizard* (Evan Miller Software) to test for statistically significant differences and correlations between the variables.

4. CHAPTER 04 – RESULTS

The sole aim of this chapter is to present the results of the statistical tests. The chapter is structured according to the primary objectives of the study, comparing different dimensions of task load placed on learners undertaking a cognitive learning activity. To recap, the experimental group were m-learners using smartphones and tablets, the control group were non-interactive learners studying a labelled textbook-style photograph.

The first section looks at the data and offers some explanations as to why there may have been deviations from a normal distribution in some of the results. The following sections present the results for each hypothesis in sequential order. The chapter concludes with summary tables of all statistically significant results and a précis of the main findings. To avoid repetition, there is no accompanying evaluation presented in this chapter, as there is an in-depth analysis and discussion presented separately in chapter 5.

4.1. DISTRIBUTION OF DATA

Statistical tests must be appropriate to the research question, but must also be appropriate to the nature of the data collected. Parametric tests such as the *t*-test and the Analysis of Variance (ANOVA) typically assume that the data follow a normal distribution whereby a majority of the data points cluster around the mean, and the extreme values lie symmetrically on either side, forming a Gaussian (bell-shaped) curve. Laerd Statistics (2016, p.12) state that a *t*-test is “fairly robust to deviations from normality” especially if the sample sizes are not too small and are equal in size – which was the case for most of the tests conducted in this study. A Shapiro-Wilk’s test was employed to determine normality of distribution and is considered to be the most powerful test for this purpose (Razali and Wah, 2011).

A Quantile-Quantile (Q-Q) plot such as shown in

Figure 4-1 and Figure 4-2 can be used to graphically represent the normality of data. The data points are created by plotting values from the sample against the values that would be expected from a standard normal distribution having the same sample size. A normally distributed sample would, therefore, follow a straight line as shown by the hashed line on

the plot. Extreme high values will tend to lie further above the line, extreme low values below.

Figure 4-1 and Figure 4-2 show that the weighted data from the NASA Task Load Index (TLX) follow a normal distribution for both groups. This was not the case for the raw data, which were non-normal for the experimental group. Hart and Staveland (2006) suggest that NASA TLX data can be used as weighted, or raw. With this in mind, the raw data were analysed to justify inclusion or exclusion in this study.

4.1.1. Distribution of NASA TLX Weighted Scores Q-Q Plots

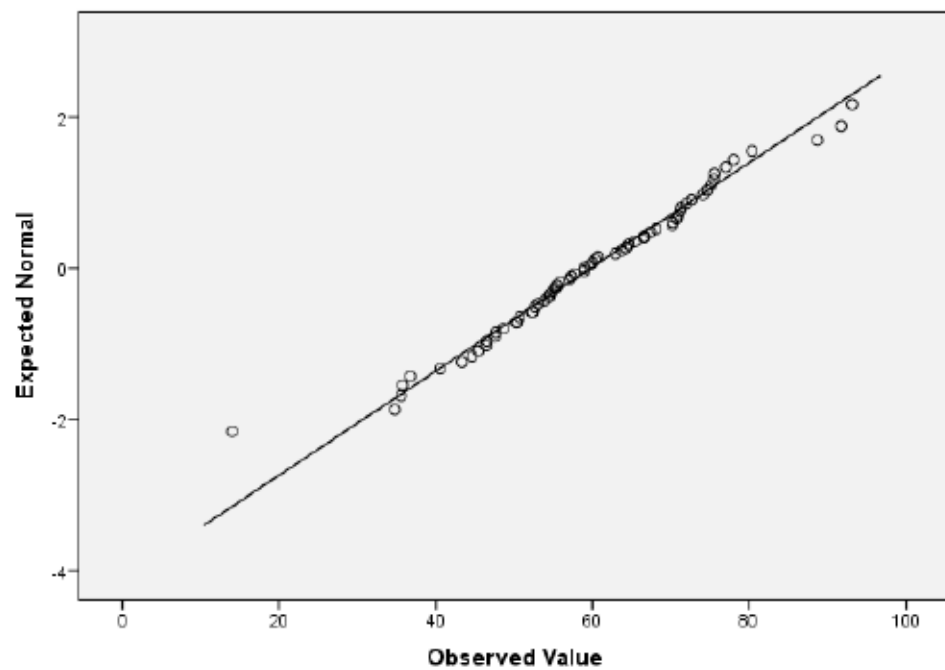


Figure 4-1: Normal distribution of the weighted NASA TLX data from the experimental group

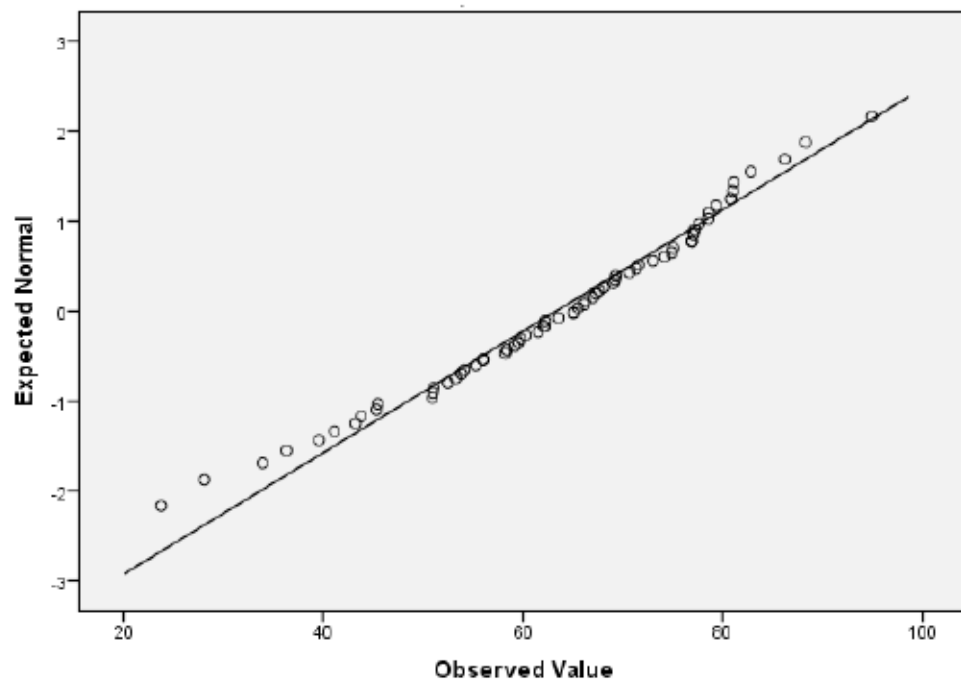


Figure 4-2: Normal distribution of the weighted NASA TLX data from the Control Group

Analysis of the Causes of Extreme-Values or Non-normality in the NASA TLX Data

Looking at the distribution plot for the raw (i.e. non-weighted) data from the experimental group, it appeared that there were some extreme values especially at the upper end of the scale. These can be seen in Figure 4-3. Outliers are extreme scores that can affect the mean value of the dataset. Moore and McCabe (1999) define outliers as data-points that fall more than $\times 1.5$ (or $\times 3.0$) the interquartile range above or below the first and third quartiles depending on whether they are moderate outliers ($\times 1.5$) or extreme outliers ($\times 3.0$). There are conflicting views on whether outliers should be included or excluded from the data before statistical analysis. Donnelly (2007) states that outliers cause unwanted distortions in statistical results and should, therefore, be excluded (trimmed from the data). Salkind (2008) argues that distortion of the results can be corrected by employing a test that uses the median, rather than the mean, as the median is not affected by the presence of extreme data-points. Exclusion is usually justified if the anomalous value is demonstrably due to experimental error, but if all of the experimental conditions have been met, the value is more likely to be a true reflection of the information being sought. The data were evaluated to determine whether there were any extreme outliers, and if so, whether data trimming was justifiable.

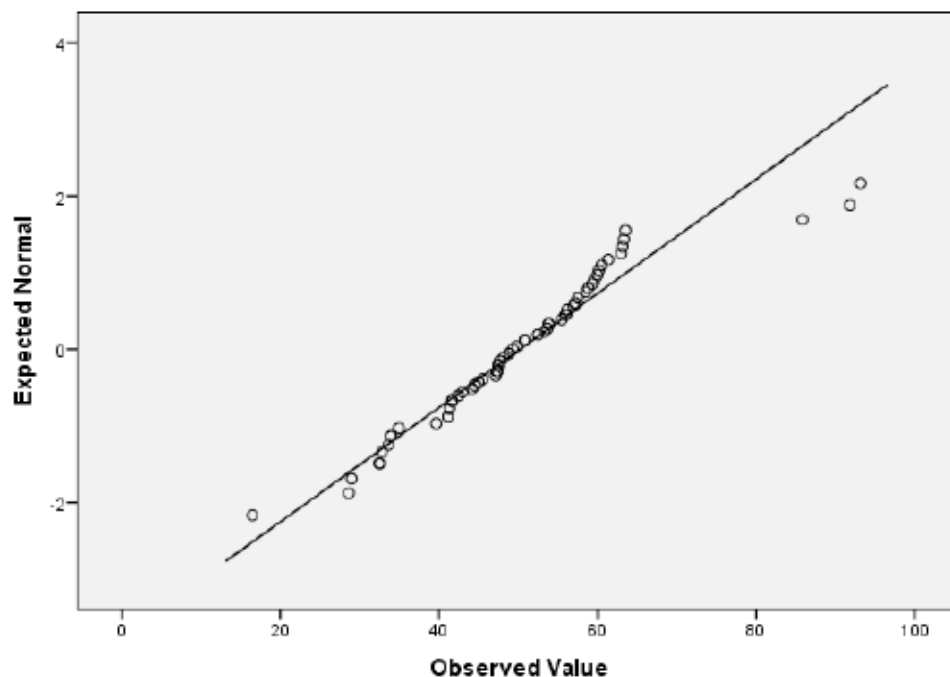


Figure 4-3: Non-normal distribution of raw NASA TLX data in the experimental group

The three core datasets (NASA TLX raw, NASA TLX weighted and Pre/Post Test score) were tested for extreme outliers by finding the interquartile range for each and calculating upper and lower fences beyond which, any data-points would fall under the definition above. There were no extreme outliers found in the weighted NASA TLX data or in any of the data for the subscales except the physical demand scale. Here it was noted that there were eleven outliers three of which were identified as being extreme outliers. These are shown in the boxplot Figure 4-4 where the box represents the second and third quartiles (50% of the data points) each side of the median and the whiskers represent the fences where data would fall more than $\times 1.5$ (or $\times 3.0$) the interquartile range above or below the first and third quartiles. The lower fence is at zero because this was the lowest score possible in the NASA TLX. The three extreme outliers fall more than $\times 3$ the interquartile range and are shown by asterisks (participants 56, 57 and 111).

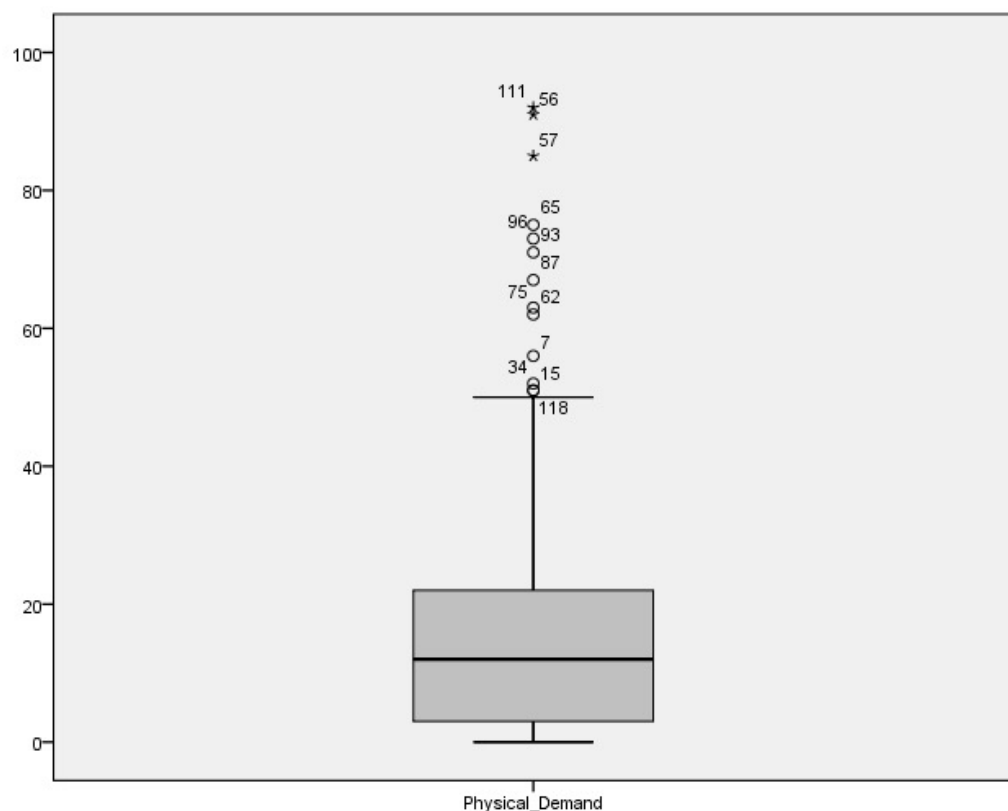


Figure 4-4: Boxplot showing outliers in the physical demand data (Y axis). See text for full explanation

The high number of outliers may be explained by the non-normal distribution of the data, namely that the data points cluster near the zero value because using a smartphone or

tablet is not particularly physically demanding. Physical demand measurement is covered in more detail in the section on *Distribution of Data from the Subscales of the NASA TLX* on page 167.

The extreme outliers were evaluated to determine whether they might have been due to experimental error. This evaluation was done with caution because the NASA TLX relies on personal perception of an event, and the scores are likely to have been a genuine reflection of the participants' perception of the task, rather than an anomaly. On checking the data there appeared to be no experimental error due to data transcription, however, as extreme values can affect the mean (and therefore limit the analysis to non-parametric testing) it was decided to look for a possible explanation for these values. The three highest scoring participants (shown by the asterisks) had given high scores to all of the test dimensions, including very high physical demand ratings (85/100, 91/100 and 92/100). Taken on a scale reflecting the difference between the lowest possible physical demand associated with any task (0) and the highest physical demand that could be associated with any task (100) these scores appear unjustifiably-high for the reasons already outlined in the instrumentation section on page 117.

All three participants also reported a very high rating for *frustration*, which may reflect issues they were having with their mobile devices. There was anecdotal evidence provided by some of the participants relating to poor network connection, and issues with their sound, in which case the frustration-rating may be justified.

These discrepancies can be assessed by looking at the NASA TLX weightings given in the pair-wise comparisons provided by the participants. Here each dimension is compared to every other dimension, and the participant is required to circle the element from each pair that they considered having contributed more to the task load. The raw score for each dimension is then weighted according to how many times that specific dimension was identified as a load-contributor in the pair-wise test ((raw score x weighting)/15). The weighted score offers a different interpretation of the individual loads in this case, for example, the participant who reported a physical demand of 62/100 as a raw score, failed to identify physical demand as being a task load contributor in the pair-wise comparisons. The participant who reported a 99% (poor) performance only identified performance as a contributing factor once (1/5) in the pair-wise comparisons. Conversely all three participants reinforced their view that the task was frustrating in the weighted scores.

The raw data from each of the dimensions reported by the lowest-scoring outlier also did not entirely agree with the data after weighting had been applied. In this case a (relatively) high raw-frustration score was reduced by the absence of this dimension being identified in any of the pair-wise comparisons.

The weighted data appeared to reflect more accurately the true opinions of the participants and a Shapiro-Wilk test demonstrated that this data followed a normal distribution. This analysis justified the use of weighted data as intended by Hart and Staveland (1988) and reaffirmed the decision to use weighted data, rather than raw data, in this study.

4.1.2. Distribution of Pre/Post Test Scores (Marks),

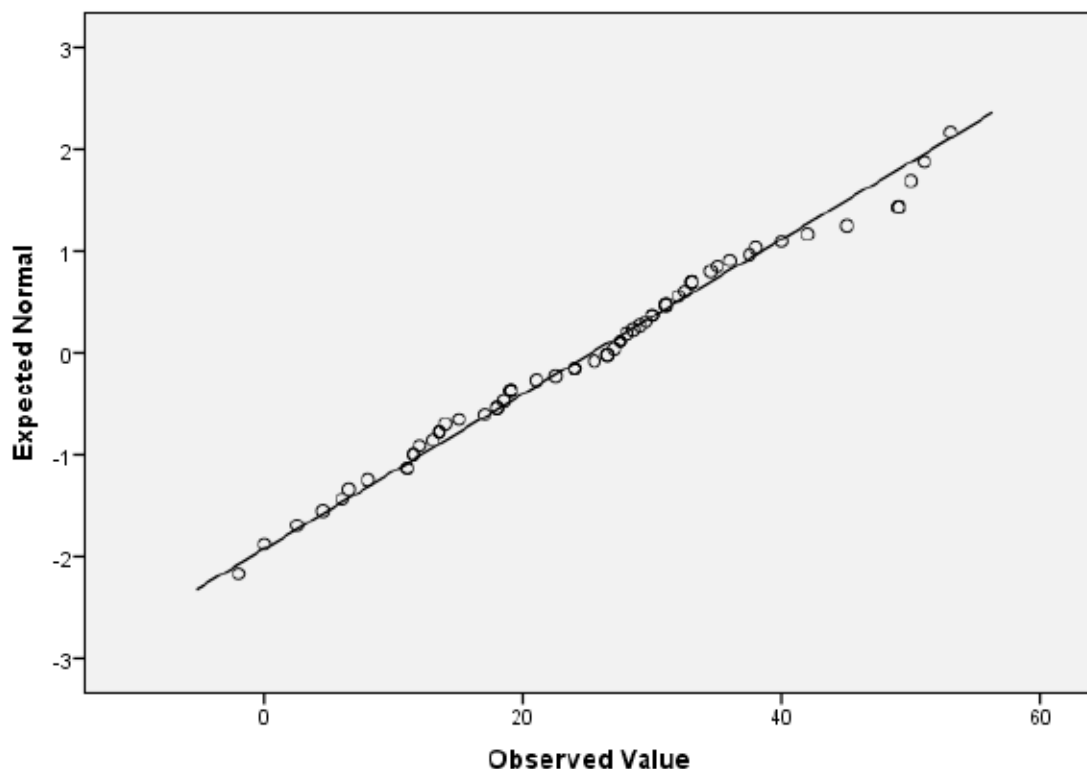


Figure 4-5: (Normal) distribution of data in the net pre/post-test score for the experimental group

Figure 4-5 shows that the data were normally distributed for the experimental group. Salkind (2008) explains that this is because there will be fewer participants who achieve

extremely high or low scores (marks), compared to the majority of participants who will achieve a mark closer to the mean.

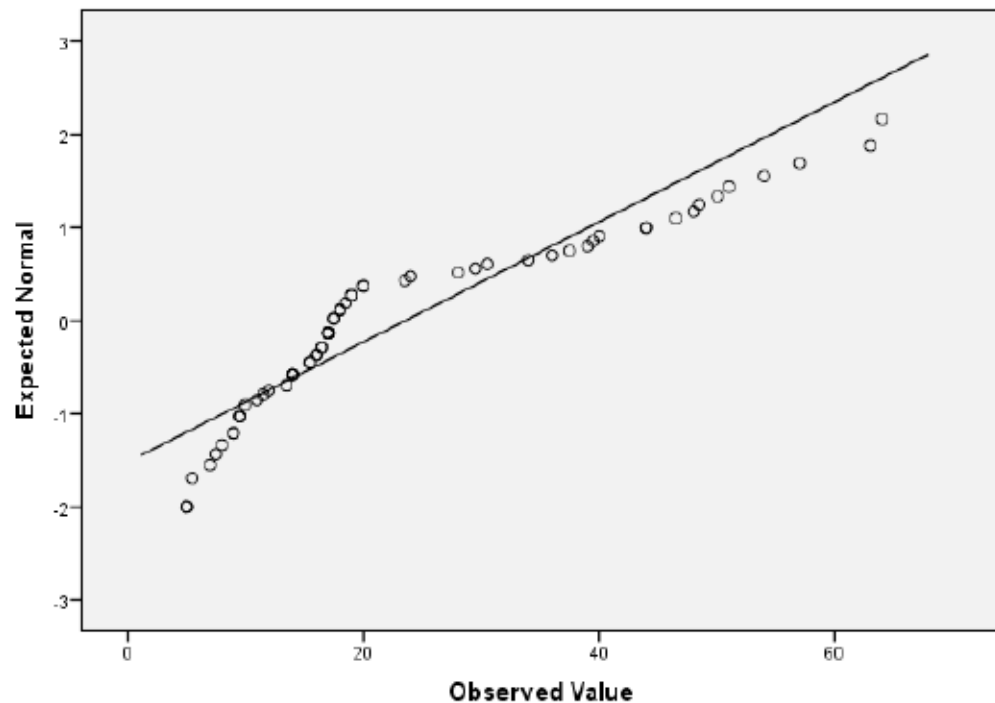


Figure 4-6: (Non-normal) distribution of data in the net pre/post-test score for the control group

Figure 4-6 shows that the data were not normally distributed for the control group. Typical causes of non-normality include; a small sample size, the presence of outliers (perhaps due to experimental error), poor measurement resolution, where there is more than one process affecting the result or where there is a boundary (such as a zero score) that prevents a symmetrical distribution of data. The sample size used in this case was large ($n=65$). The marks were checked, and there were no experimental errors evident. The measurement resolution was high with a range of possible marks between 0-117. The experimental conditions were the same for all of the control-group participants so it is unlikely that more than one process or variable could have affected the result. (Interestingly, this was not the case for the experimental group, who were using mobile devices from different manufacturers, yet yielded results that demonstrated a normal distribution). A closer analysis of the data revealed that there were no extreme outliers,

The reason for the non-normality, therefore, seems to be because the inherent difficulty of the test resulted in the mean marks for both groups being closer to zero than to the maximum possible mark (117). The mean mark for the control group was lower than for the experimental group, and this has positively skewed the data towards the zero point, the negative tail being truncated by this boundary (see Figure 4-7).

Although zero seems to be the logical lower boundary in a test of this nature, there was one participant in the experimental group who made some correct guesses in the pre-test that were not replicated in the post-test resulting in a minus score (-2).

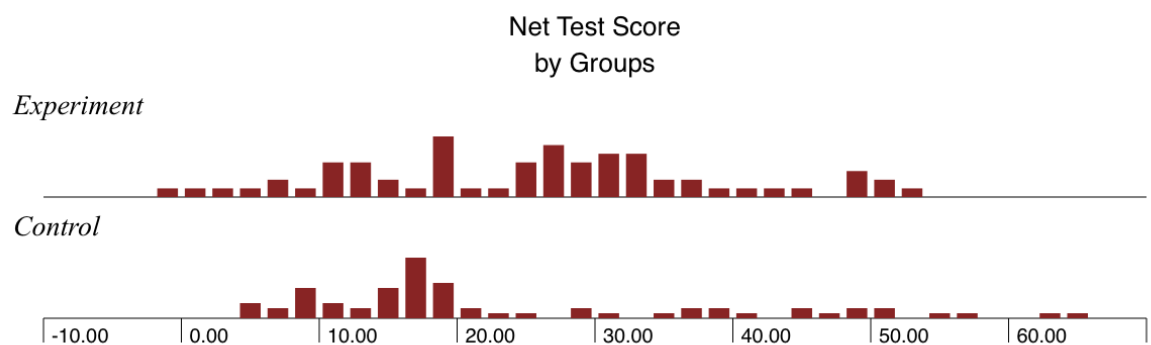


Figure 4-7: Histogram showing comparison of the distribution of pre/post-test marks between groups

4.1.3. Distribution of Data from the Subscales of the NASA TLX

In looking at the separate dimensions of the NASA TLX test, there were other data that did not follow a normal distribution, these were:

Mental demand – control group

Physical demand – both groups

Effort – control group

Frustration – control group

A possible explanation for these non-normal distributions is that the data for each dimension are raw (non-weighted) data. The possible reasons for non-normality were covered on page 162 (Analysis of the Causes of Extreme-Values or Non-normality in the NASA TLX Data). In addition, the physical demand subscale values clustered around zero truncating the negative tail of the curve.

The collected data were transcribed from the paper collection sheets into an Excel (Microsoft Corp.) workbook. The data were sorted into columns relating to:

Participant information

ID number

gender

age in years

Device used (for experimental group)

phone or tablet

screen size

manufacturer

pixel density (spatial resolution of the display)

Pre and post-test results

answers given

marks awarded

NASA TLX ratings

net score (weighted and raw)

mental demand,

physical demand,

temporal demand,

performance,

effort,

frustration

4.2. SOFTWARE USED FOR STATISTICAL ANALYSIS

The workbook data were imported into two data analysis packages, *SPSS* (Statistical Package for the Social Sciences) (International Business Machines Corp.) and *Wizard* (Evan Miller Software). The rationale for seeking agreement between two different data-analysis tools was to ensure reliability and consistency of data processing. The two packages gave identical results in all tests conducted.

4.3. MARKING PRE AND POST-TEST SCORES

All of the pre-tests and post-tests were marked and re-marked until no further marking errors were discovered. Four iterations were required. This method was used in preference to using a second marker because it reduced any possibility that both markers might make similar errors in marking and eliminated any bias that might have been due to the marking style of individual markers. Marking of the pre/post test scores was performed blind to reduce experimenter-bias.

Two types of marking criteria were considered:

Strict Marking: whereby marks were only awarded where correctly spelt and correctly structured, fully complete answers were provided.

Lenient Marking: whereby marks were awarded for partially correct answers. These included incomplete answers (for example the answer “carotid” where the full answer was “carotid canal”).

Both sets of marks were analysed using the statistical data analysis software. There was no statistically significant difference in outcomes across any of the NASA sub-scales. In the context of the study, it was thought to be more appropriate to use the lenient marking criteria rather than the strict criteria. This was for two reasons. Firstly, the wider range of marks provided a more sensitive measure for assessing the memorisation of anatomical structures. For example, if the correct answer was “facial nerve” and the participant answered “mandibular nerve” one mark would be awarded for identifying that a nerve passed through the foramen. Similarly, a mark would be awarded for correctly recalling the name of a structure, but not the type of structure (e.g. artery or nerve).

The second reason for choosing a lenient marking scheme was that it provided better measurement resolution, which is a factor that affects the normality of distribution. The maximum achievable score for correctly naming all of the structures was 117 marks using the lenient marking scheme, compared to 36 marks using the strict criteria.

4.3.1. Marking Criteria for Pre and Post-Testing

- All words correctly remembered (including where used inappropriately e.g. “Jugular Artery” instead of “Jugular Vein”) **1 mark**
- Alternative Latin or non-UK spelling or translations **1 mark**.
- Words phonetically correct but misspelt **0.5 mark**
- No marks were awarded for identifying the word “foramen” as this was given in the question
- To avoid measurement bias, extra marks were awarded for pairing the correct anatomical name to the correct structure. For example, the *incisive foramen* transmits the *nasopalatine nerves* and the *sphenopalatine artery*. If a participant answered with an incorrect pairing (nasopalatine artery, sphenopalatine nerves) a score of 4 was awarded to acknowledge the fact that they had learnt that the foramen transmits a nerve (**1 mark**) an artery (**1 mark**) that one of the structures was called nasopalatine (**1 mark**) and that the other structure was called sphenopalatine (**1 mark**). To differentiate this score from a participant who paired the terms correctly, additional marks were awarded for correct pairing (nasopalatine nerve, sphenopalatine artery) giving a score of **6 marks**. An alternative marking strategy would have been to assign a zero score for an incorrect pairing (nasopalatine artery). This would have also have introduced measurement bias because an answer that contained the single word *artery* would receive a score of 1 (for correctly identifying that an artery passed through the foramen).

4.4. PARTICIPANT DEMOGRAPHICS

4.4.1. Experimental Group

Interactive m-learning activity group (n=65)

Experimental Group Gender Ratio

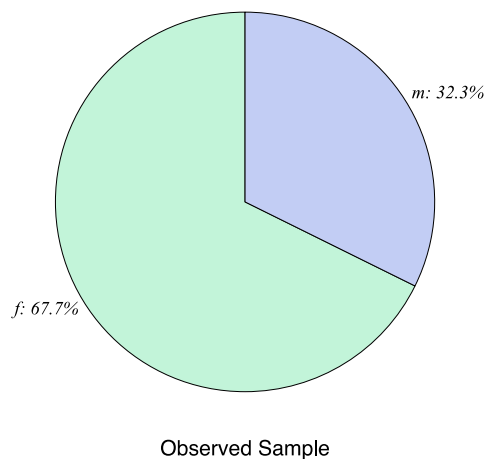
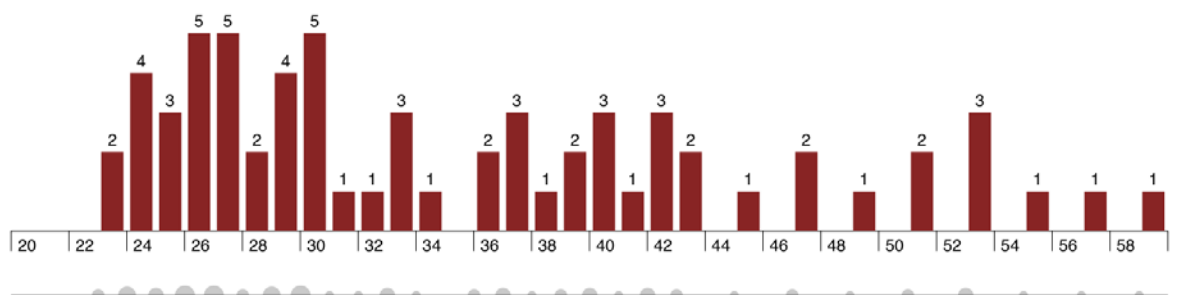


Figure 4-8: Pie chart showing experimental group gender ratio

Experimental Group Age Distribution

Distribution of Age



Estimated mean = 35.262 ± 2.432

Figure 4-9: Histogram showing experimental group age distribution and mean

The experimental group age range was 23 years to 59 years, the mean age was 35.26 years.

4.4.2. Control Group

Non-interactive learning activity group (n=65)

Control Group Gender Ratio

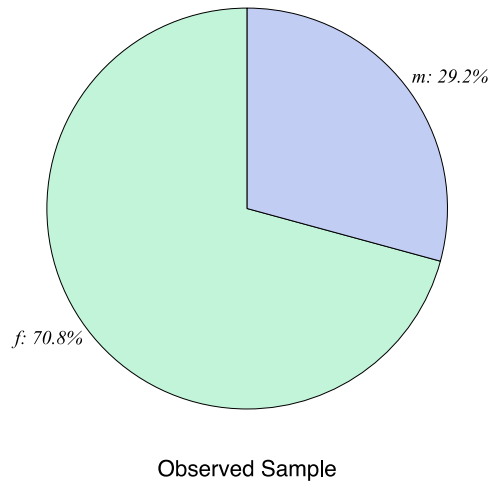


Figure 4-10: Pie chart showing control group gender ratio

Control Group Age Distribution

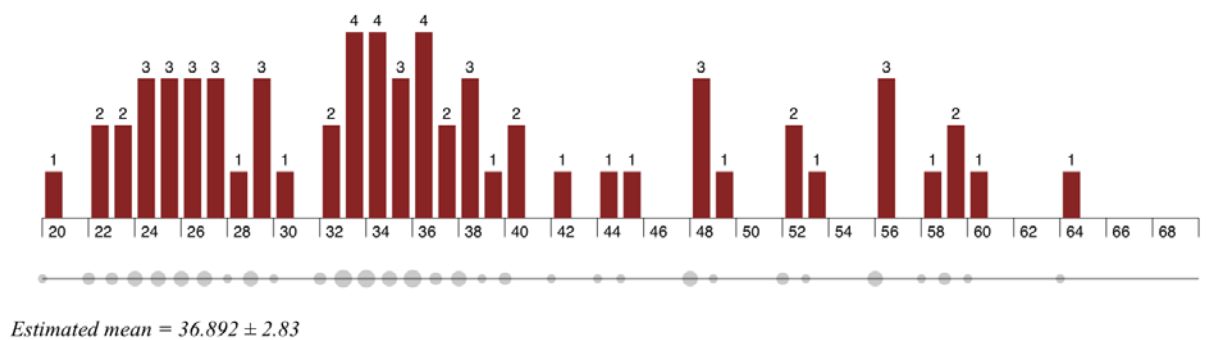


Figure 4-11: Histogram showing control group age distribution and mean

The control group age range was 20 years to 64 years, the mean was 36.89 years.

4.5. PRINCIPAL FINDINGS

4.5.1. Comparing Net Task load Between Groups (H1)

Null Hypothesis:

H1: There will be no statistically significant difference in NASA TLX task load score between the experimental group (mobile-learners) and the control group (non-interactive learners).

Result

An independent-samples t-test was run to determine if there were differences in NASA TLX task load score between mobile-learners and non-interactive learners. There were no extreme outliers in the data, as assessed by inspection of a boxplot. Task load scores for each group were normally distributed, as assessed by Shapiro-Wilk's test ($p > .05$), and there was homogeneity of variances, as assessed by Levene's test for equality of variances ($p=.824$). The test indicated no statistically significant difference between the mobile-learners group ($M=59.81 \pm 3.59$) and the non-interactive group ($M=63.32 \pm 3.66$), $t(128)=1.36$, $p=.823$.

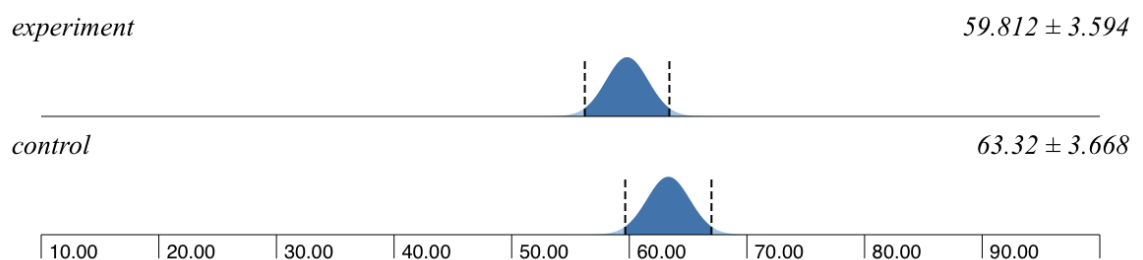


Figure 4-12: NASA TLX (mean) net task load score between groups

Group		N	Mean	Std. Deviation	Std. Error Mean
NASA TLX	Mobile-Learners	65	59.81	14.50	1.79
Weighted Score	Non-Interactive	65	63.32	14.80	1.83

Table 4-1 Group Statistics for NASA TLX Score between groups

Conclusion

There was no statistically significant difference between group means. The null hypothesis cannot be rejected.

4.5.2. Comparing Test Scores Between Groups (H2)

Null hypothesis

H2: There will be no difference in pre/post-test scores between the experimental group and the control group.

The analysis relating to pre/post test scores used two sets of data. One set featured lenient marking criteria and the other used strict marking criteria.

Lenient Marking Result

The pre/post-test score data were not normally distributed for the control group. A Mann-Whitney U test was run to determine if there were differences in pre/post-test score between mobile-learners and non-interactive learners. Distributions of the pre/post-test scores for mobile-learners and non-interactive learners were not similar, as assessed by visual inspection. Median pre/post-test score for mobile-learners (mean rank=70.30) and non-interactive learners (mean rank=60.70) was not statistically significantly different, $U=2,424.50$, $z=1.45$, $p=.146$

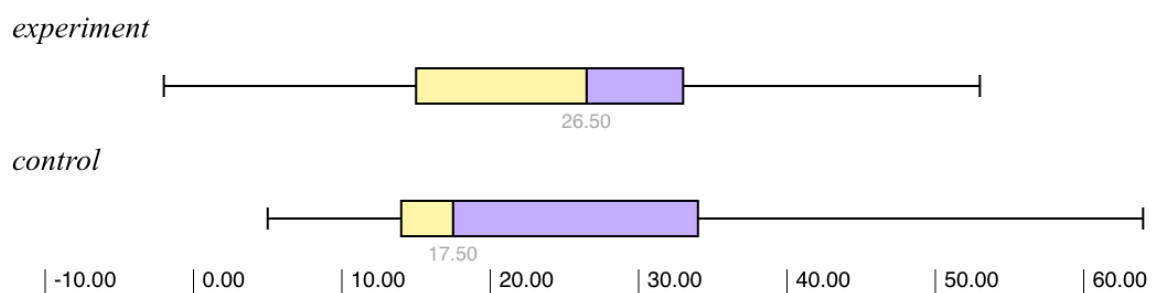


Figure 4-13: Boxplot of net test-score between groups (lenient marking scheme) (median values shown)

Conclusion

There was no statistically significant difference between mean rank pre/post-test scores between the groups. The null hypothesis cannot be rejected.

Strict Marking Result

The pre/post-test score data were not normally distributed for the control group. A Mann-Whitney U test was run to determine if there were differences in pre/post-test score between mobile-learners and non-interactive learners. Distributions of the pre/post-test scores for mobile-learners and non-interactive learners were not similar, as assessed by visual inspection. Median pre/post-test score for mobile-learners (mean rank=70.15) and non-interactive learners (mean rank=60.85) was not statistically significantly different, $U=1,810.00$, $z=1.413$, $p=.158$

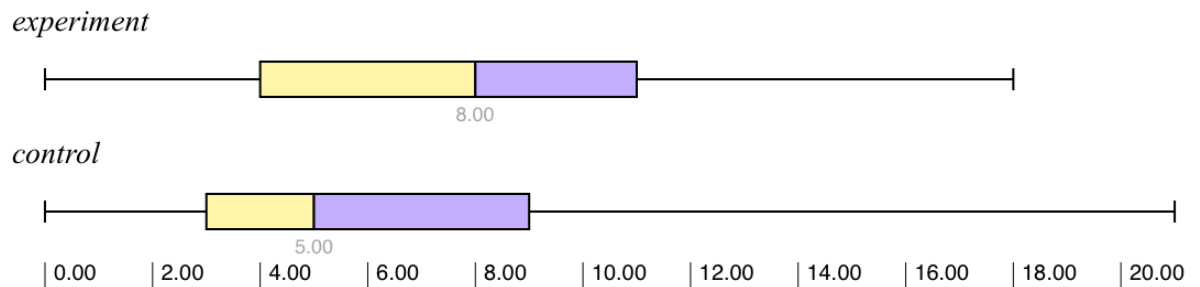


Figure 4-14: Boxplot of net test-score between groups (strict marking scheme) (median values shown)

Conclusion

There was no statistically significant difference between mean ranks. The null hypothesis cannot be rejected.

4.5.3. Comparing Mental Demand Between Groups (H3)

Null Hypothesis:

H3: There will be no statistically significant difference in reported mental demand between the experimental group and the control group.

Mental Demand Descriptor

How mentally demanding was the learning task?

Result

The data did not follow a normal distribution for either group.

A Mann-Whitney U test was run to determine if there were differences in mental demand between mobile-learners and non-interactive learners. Distributions of mental demand for mobile-learners and non-interactive learners were similar, as assessed by visual inspection. Median mental demand was statistically significantly higher in non-interactive learners (76.00) than in mobile-learners (70.00), $U=1,589.00$, $z=2.44$, $p=.015$, $r=.214$.

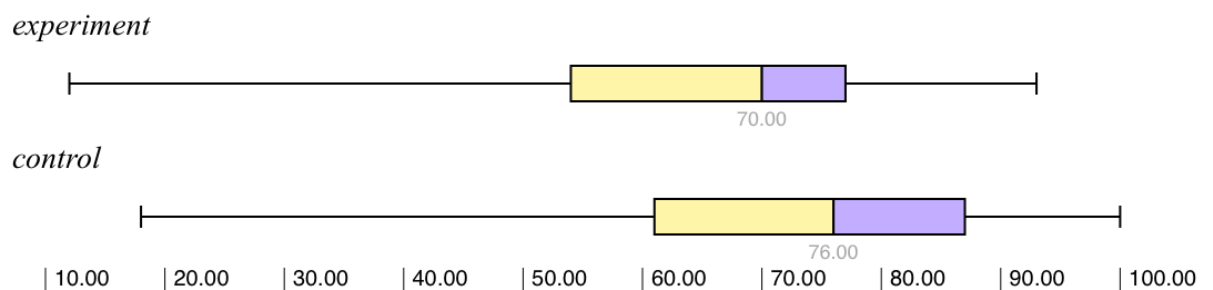


Figure 4-15: Boxplot of reported levels of mental demand between groups (median values shown)

Conclusion

The null hypothesis can be rejected. There is a statistically significant difference in median mental-demand score between the two groups. The non-interactive learners reported a higher level of mental demand during the learning activity than the mobile-learners. The effect size was small.

4.5.4. Comparing Physical Demand Between Groups (H4)

Null Hypothesis:

H4: There will be no statistically significant difference in physical demand reported between the experimental group and the control group.

Physical Demand Descriptor

How physically demanding was the learning task?

Result

The data were not normally distributed. There were two extreme outliers in the experimental group (mobile-learners) (scores 85/100, 91/100), and one extreme outlier in the control group (non-interactive learners) (92/100).

A Mann-Whitney U test was run to determine if there were differences in physical demand between mobile-learners and non-interactive learners. Distributions of physical demand for mobile-learners and non-interactive learners were similar, as assessed by visual inspection. Median reported physical demand scores for mobile-learners (14.00) and non-interactive learners (8.00) were not statistically significantly different, $U=2,406.5$, $z=1.37$, $p=.170$.

Excluding the extreme outliers did not change the significance of the result

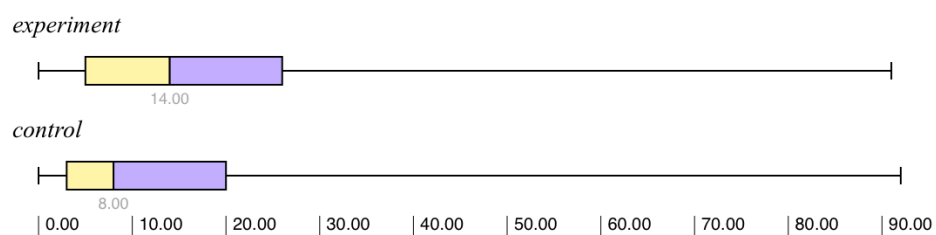


Figure 4-16: Boxplot of reported levels of physical demand between groups (median values shown)

Conclusion

There was no statistically significant difference in median physical-demand score between groups. The null hypothesis cannot be rejected.

4.5.5. Comparing Temporal Demand Between Groups (H5)

Null Hypothesis:

H5: There will be no statistically significant difference in temporal demand reported between the experimental group and the control group.

Temporal Demand Descriptor

How much time pressure did you feel due to the rate of pace at which the tasks or task elements occurred?

Result

An independent-samples t-test was run to determine if there were differences in temporal demand between mobile-learners and non-interactive learners. There were no extreme outliers in the data, as assessed by inspection of a boxplot. Temporal demand scores for each group were normally distributed, as assessed by Shapiro-Wilk's test ($p > .05$), and there was homogeneity of variances, as assessed by Levene's test for equality of variances ($p=.167$). The test indicated no statistically significant difference between the mobile-learners group ($M=61.86 \pm 4.65$) and the non-interactive group ($M=56.03 \pm 5.62$), $t(128)=1.598$, $p=.112$.

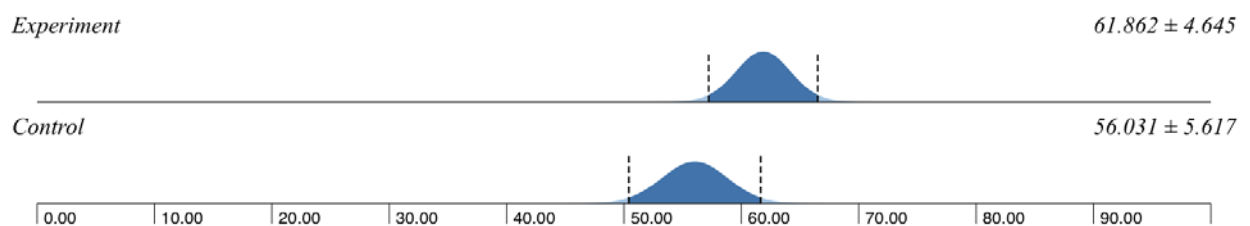


Figure 4-17: Reported levels of temporal demand between groups

Group Statistics

Group		N	Mean	Std. Deviation	Std. Error Mean
Temporal	Mobile-Learners	65	61.86	18.75	2.32
Demand	Non-Interactive	65	56.03	22.67	2.81

Table 4-2 Group statistics for NASA TLX score between groups

Conclusion

There was no statistically significant difference between group means. The null hypothesis cannot be rejected.

4.5.6. Comparing Performance Between Groups (H6)

Null Hypothesis:

H6: There will be no statistically significant difference in reported task-related performance between the experimental group and the control group.

Performance Descriptor

How successful were you in accomplishing what you were supposed to do?

Result

The control group data were not normally distributed.

A Mann-Whitney U test was run to determine if there were differences in performance between mobile-learners and non-interactive learners. Distributions of the performance scores for mobile-learners and non-interactive learners were not similar, as assessed by visual inspection. (Poor) performance scores for the non-interactive group (mean rank=71.99) were statistically significantly higher than for the mobile-learners (mean rank=59.01), $U=2,534.5$, $z=-1.966$, $p=.049$, $r=.172$.

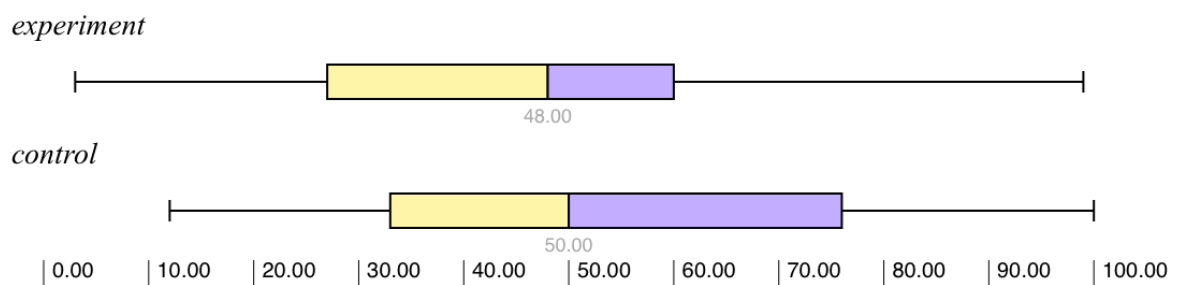


Figure 4-18: Boxplot of reported level of performance between groups (median values shown)

Conclusion

NASA Descriptors; 0=perfect performance, 100=failure. Higher scores =poorer perceived performance (perceived failure).

The null hypothesis can be rejected. There is a statistically significant difference in mean-rank performance score between the two groups. The non-interactive learners reported a higher level of poor performance during the learning-activity than the mobile-learners. The effect size was small.

Reported performance was also tested against pre/post test score to see if there was any correlation to support the use of pre/post-testing as a performance outcome measure.

This is an indirect objective method of assessing cognitive load.

A Spearman's rank-order correlation was run to assess the relationship between reported performance and pre/post test score results. Preliminary analysis showed the relationship to be monotonic, as assessed by visual inspection of a scatter plot.

There was a weak but highly significant negative correlation between reported performance and pre/post-test score, $r_s(130)=-.191, p=.002$.

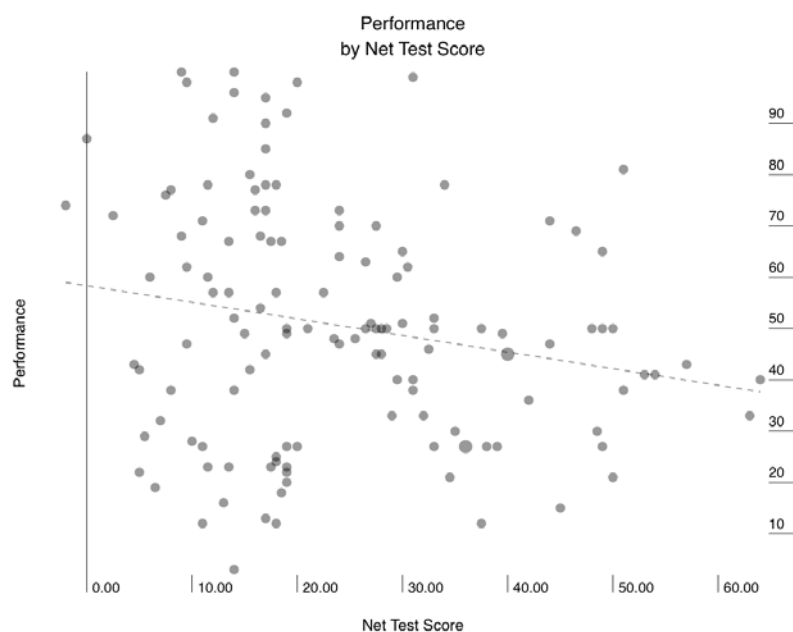


Figure 4-19: Scatterplot showing weak negative correlation between performance and net test score

4.5.7. Comparing Effort Between Groups (H7)

Null Hypothesis:

H7: There will be no statistically significant difference in reported effort between the experimental group and the control group.

Effort Descriptor

How hard did you have to work to accomplish your level of performance?

Result

The control group data were not normally distributed.

A Mann-Whitney U test was run to determine if there were differences in effort between mobile-learners and non-interactive learners. Distributions of effort for mobile-learners and non-interactive learners were similar, as assessed by visual inspection. Median reported effort score for mobile-learners (65.00) and non-interactive learners (72.00) was not statistically significantly different, $U=2,438$, $z=1.52$, $p=.129$

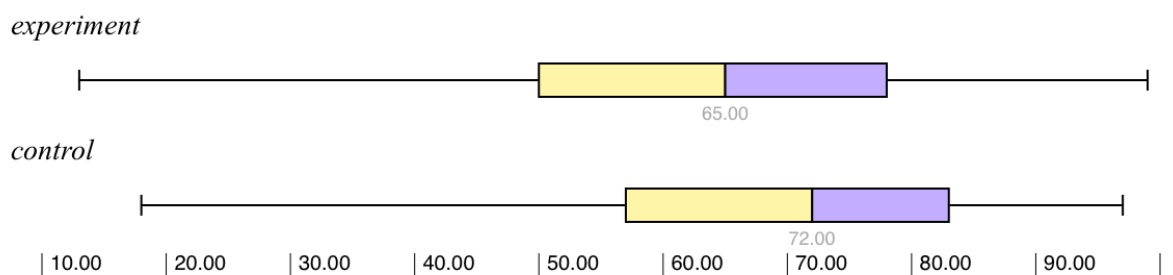


Figure 4-20: Boxplot of reported level of effort between groups (median values shown)

Conclusion

There is no statistically significant difference in median effort score between the two groups. The null hypothesis cannot be rejected.

Reported effort was also tested against mental demand and physical demand to determine which of these dimensions contributed to the overall effort score. The rationale for this test is provided in the following chapter in section 5.2.9.

A Spearman's rank-order correlation was run to assess the relationship between reported effort and physical demand. Preliminary analysis showed the relationship to be monotonic, as assessed by visual inspection of a scatterplot.

There was no statistically significant correlation between reported physical demand and effort, $r_s(130)=.089, p=.315$.

A Spearman's rank-order correlation was run to assess the relationship between reported effort and mental demand. Preliminary analysis showed the relationship to be monotonic, as assessed by visual inspection of a scatterplot.

There was a moderate, statistically significant positive correlation between reported mental demand and effort, $r_s(130)=.578, p < .001$.

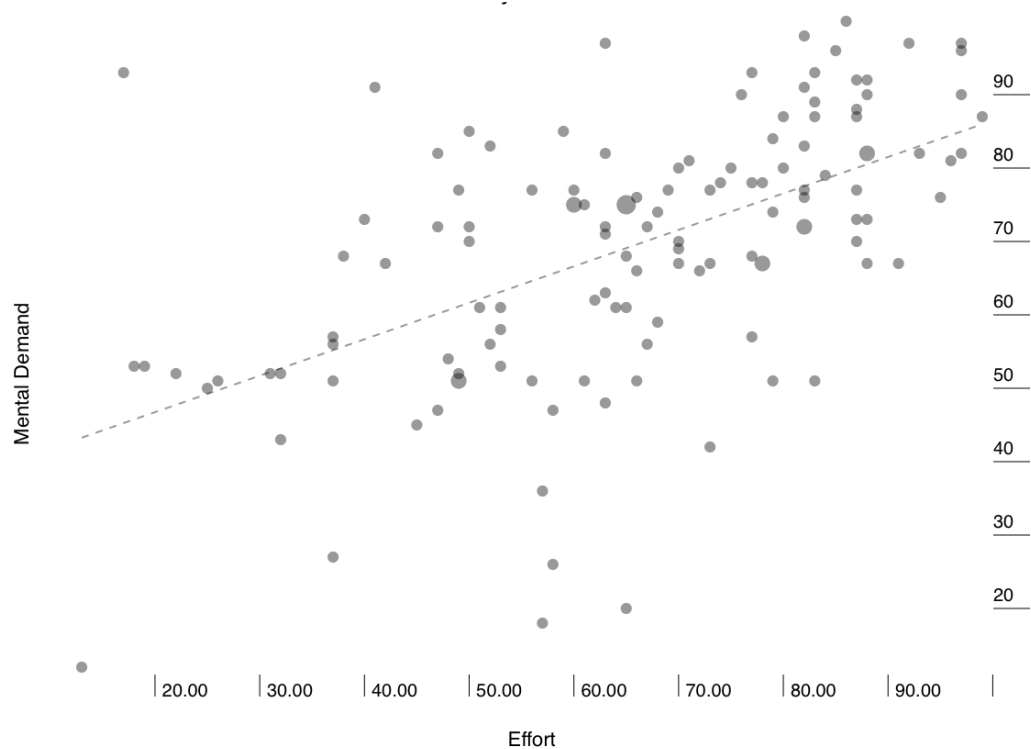


Figure 4-21: Scatterplot showing a medium positive correlation between mental demand and effort

4.5.8. Comparing Frustration Between Groups (H8)

Null Hypothesis:

H8: There will be no statistically significant difference in reported task-related frustration between the experimental group and the control group.

Frustration

How insecure, discouraged, irritated, stressed and annoyed did you feel during the task?

Result

The control group data were not normally distributed.

A Mann-Whitney U test was run to determine if there were differences in frustration score between mobile-learners and non-interactive learners. Distributions of the frustration scores for mobile-learners and non-interactive learners were not similar, as assessed by visual inspection. Median frustration score for mobile-learners (mean rank=62.15) and non-interactive learners (mean rank=68.85) was not statistically significantly different, $U=2,330$, $z=1.013$, $p=.311$

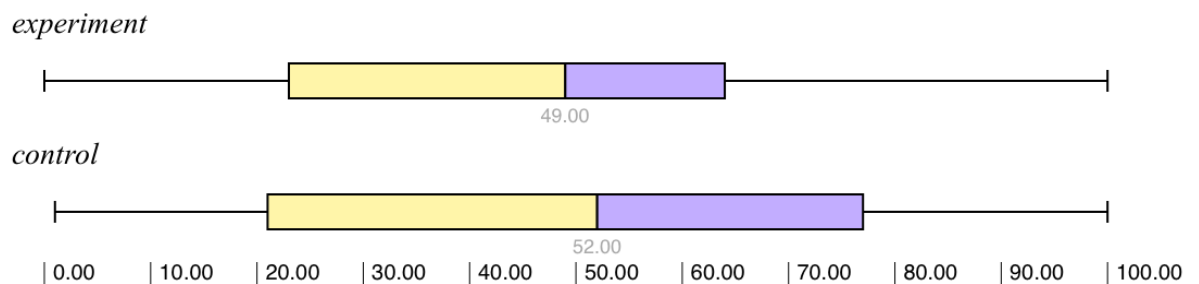


Figure 4-22: Boxplot of reported level of frustration between groups (median values shown)

Conclusion

There is no statistically significant difference in mean-rank frustration score between the two groups. The null hypothesis cannot be rejected.

4.5.9. Summary of Results for Comparisons Between Groups

Greyed-out data refers to results obtained from tests that were not optimal for the data under analysis but were undertaken for comparison purposes. Text highlighted in blue refers to data from non-parametric tests and indicates the median score rather than the mean. Asterisks indicate statistically significant results ($p < .05$).

Dependent Variables	Experimental group (n=65) score <i>M (SD)/Mdn (IQR)</i>	Control group (n=65) score <i>M (SD)/Mdn (IQR)</i>	Mann-Whitney p-value	Mann-Whitney CI (%)	t-test p-value	t-test CI (%)	Sig?	Effect size
(H2) Pre-Post Test score (%)	26.50 (18)	17.50 (21)	0.146	85	0.5	<80	No	
(H1) Net Task load	59.81 (14.51)	63.32 (14.80)	0.104	89	0.823	82	No	
(H3) Mental Demand	70.00 (23)	76.00 (26)*	0.015	98	0.045	95	Yes	$r=.214$
(H4) Physical Demand	14.00 (21)	8.00 (17)	0.170	82	0.390	<80	No	
(H5) Temporal Demand	61.86 (18.75)	56.03 (22.67)	0.196	81	0.112	88	No	
(H6) Performance	48.00 (33)	50.00 (44)*	0.049	95	0.029	97	Yes	$r=.172$
(H7) Effort	65.00 (28)	72.00 (26)	0.129	87	0.173	80	No	
(H8) Frustration	49.00 (57.5)	52.00 (42)	0.311	<80	0.341	<80	No	

Table 4-3: Summary of all comparisons between experimental group and control group

The table shows that the non-interactive learners reported a statistically significantly higher level of mental demand than the mobile-learners and the non-interactive learners reported a statistically significantly poorer perceived level of performance in the task (higher values = poorer performance).

4.6. SIGNIFICANT RESULTS BETWEEN SMARTPHONE-LEARNERS ($n=36$) AND CONTROL GROUP ($n=65$)

4.6.1. Comparing Net Task load Between Smartphone and Control Groups (H9)

Null Hypothesis:

H9: There will be no statistically significant difference in NASA TLX task load score between the experimental group (smartphone-learners) ($n=36$) and the control group (non-interactive learners)($n=65$).

Result

An independent-samples t-test was run to determine if there were differences in NASA TLX task load score between smartphone-learners and non-interactive learners. There were no extreme outliers in the data, as assessed by inspection of a boxplot. Task load scores for each group were normally distributed, as assessed by Shapiro-Wilk's test ($p > .05$), and there was homogeneity of variances, as assessed by Levene's test for equality of variances ($p=.749$). The reported net task load was greater for the non-interactive group (63.32 ± 4.07) than for the smartphone-learners group (57.14 ± 4.55), a statistically significant difference of -6.18 (95% CI, -12.09 to -0.272), $t(99)=2.076$, $p=.040$, $d=.460$.

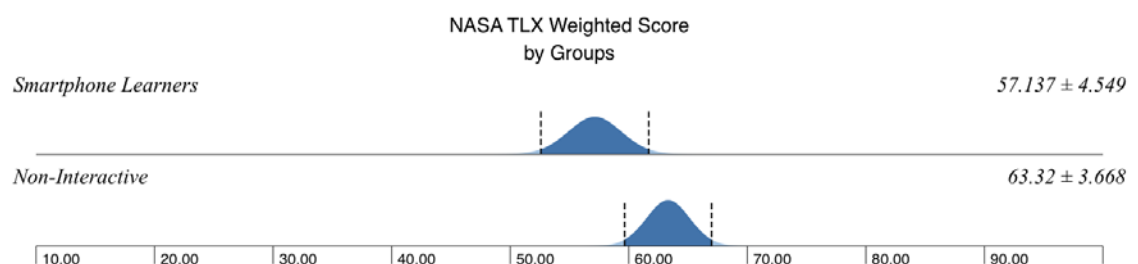


Figure 4-23: NASA TLX (mean) net task load score between non-interactive learners and smartphone-learners

Group Statistics					
Group		<i>n</i>	Mean	Std. Deviation	Std. Error Mean
NASA TLX	smartphone-Learners	36	57.14	13.44	2.24
Weighted Score	Non-Interactive	65	63.32	14.80	1.83

Table 4-4 Group Statistics for NASA TLX Score between groups

Conclusion

The null hypothesis can be rejected. The group-means were statistically significantly different. The non-interactive learners reported a higher overall task load than the smartphone-learners. The effect size was small.

4.6.2. Comparing Mental Demand Between Smartphone and Control Groups (H11)

Null Hypothesis:

H11: There will be no statistically significant difference in mental demand between the experimental group (smartphone-learners) and the control group (non-interactive learners).

Result

The data did not follow a normal distribution for either group.

A Mann-Whitney U test was run to determine if there were differences in reported mental demand between smartphone-learners and non-interactive learners. Distributions of reported mental demand for smartphone-learners and non-interactive learners were similar, as assessed by visual inspection. Median reported mental demand was statistically significantly higher in non-interactive learners (76.00) than in mobile-learners (66.00), $U=741.00$, $z=3.04$, $p=.002$, $r=.302$.

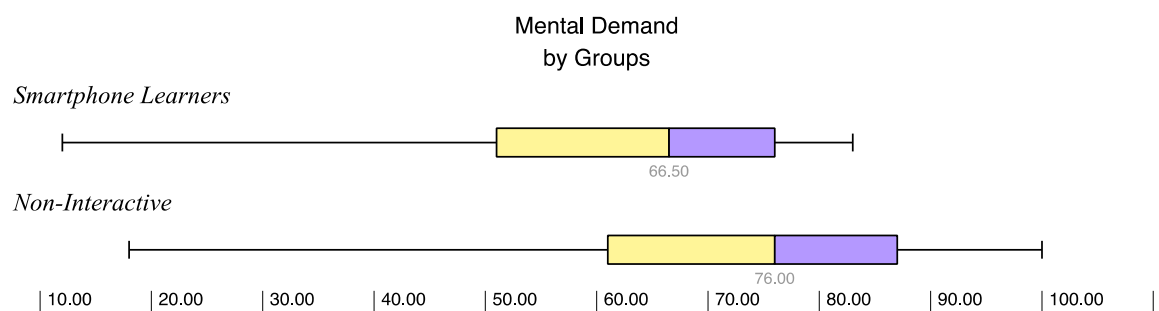


Figure 4-24: Boxplot of reported levels of mental demand between groups (median values shown)

Conclusion

The null hypothesis can be rejected. There is a statistically significant difference in median mental-demand score between the two groups. The non-interactive learners reported a

higher level of mental demand during the learning-activity than the smartphone-learners.
The effect size was medium.

4.6.3. Comparing Effort Between Smartphone and Control Groups (H15)

Null Hypothesis:

H15: There will be no statistically significant difference in effort reported between the smartphone-learners and the control group.

Result

The control group data were not normally distributed.

A Mann-Whitney U test was run to determine if there were differences in reported effort between mobile-learners and non-interactive learners. Distributions of the reported effort scores for smartphone-learners and non-interactive learners were not similar, as assessed by visual inspection. Effort scores for the non-interactive group (mean rank=59.08) were statistically significantly higher than for the mobile-learners (mean rank=43.13), $U=860$, $z=-2.199$, $p=.028$, $r=.219$.

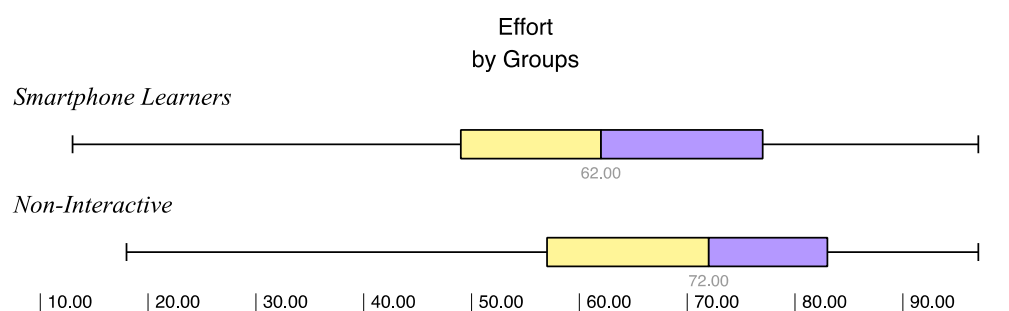


Figure 4-25: Boxplot of reported level of effort between groups (median values shown)

Conclusion

The null hypothesis can be rejected. There is a statistically significant difference in mean-rank performance score between the two groups. The non-interactive learners reported a higher level of effort during the learning-activity than the smartphone-learners. The effect size was small.

4.7. SIGNIFICANT RESULTS BETWEEN TABLET-LEARNERS ($n=24$) AND CONTROL GROUP ($n=65$)

4.7.1. Comparing Physical Demand Between Tablet and Control Groups (H20)

Null Hypothesis:

H20: There will be no statistically significant difference in physical demand reported between the tablet-learners and the control group.

Result

The data were not normally distributed. A Mann-Whitney U test was run to determine if there were differences in reported physical demand between tablet-learners and non-interactive learners. Distributions of the physical demand scores for tablet-learners and non-interactive learners were not similar, as assessed by visual inspection. Physical demand score for the tablet-learners group (mean rank=57.19) was statistically significantly higher than for the non-interactive learners (mean rank=43.18), $U=661.5$, $z=-2.304$, $p=.020$, $r=.224$.

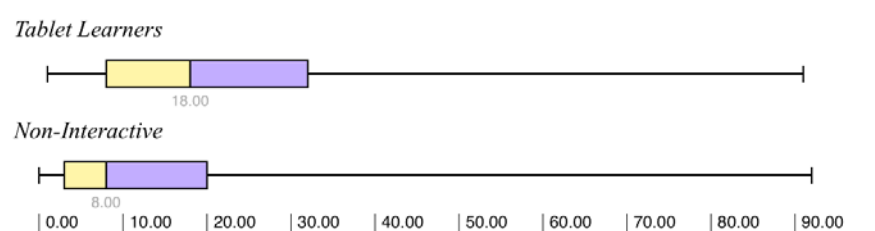


Figure 4-26: Boxplot of reported levels of physical demand between groups (median values shown)

Conclusion

The null hypothesis can be rejected. There is a statistically significant difference in median-rank physical-demand score between the two groups. The tablet-learners reported a higher level of physical demand during the learning-activity than the non-interactive learners. The effect size was small.

4.8. SECONDARY FINDINGS: RESULTS RELATING TO DEVICE TYPE

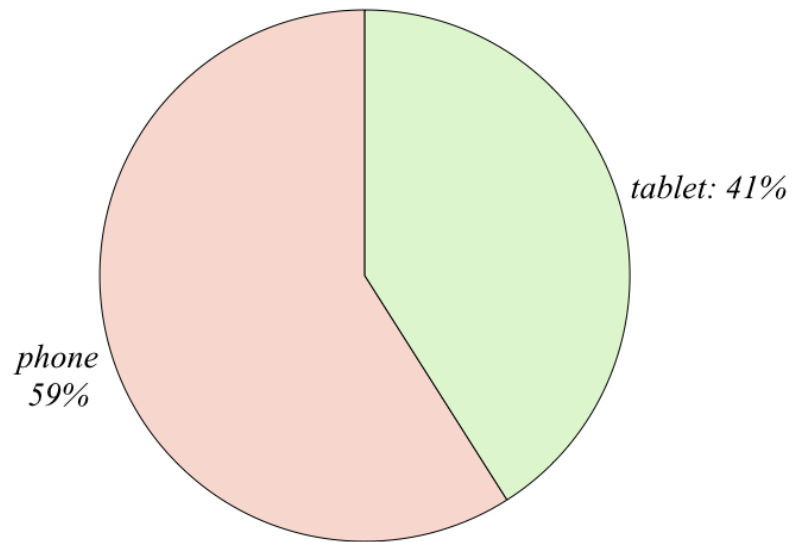


Figure 4-27: Pie chart showing distribution of device type

Device Type

Participants in the m-learning group were randomly allocated into two sub-groups, smartphone-learners and tablet-learners. Five participants neglected to identify which type of device had been used on the NASA response form.

Smartphone learners (n=36)

Tablet learners (n=24)

Not provided (n=5)

4.8.1. Comparing Net Task load Between Learners Grouped by Device Type (H25)

Null Hypothesis:

H25: There will be no statistically significant difference in NASA TLX task load score between the learners grouped by device type (smartphone vs. tablet computer).

Result

An independent-samples t-test was run to determine if there were differences in NASA TLX task load score between smartphone-learners and tablet-learners. There were no extreme outliers in the data, as assessed by inspection of a boxplot. Task load scores for each group were normally distributed, as assessed by Shapiro-Wilk's test ($p > .05$), and there was homogeneity of variances, as assessed by Levene's test for equality of variances ($p = .371$). The reported net task load was greater for the tablet-learners group (64.75 ± 6.30) than for the smartphone-learners group (57.14 ± 4.55), a statistically significant difference of -7.61 (95% CI, -0.20 to -15.02), $t(58) = 2.06$, $p = .044$, $d = .566$.

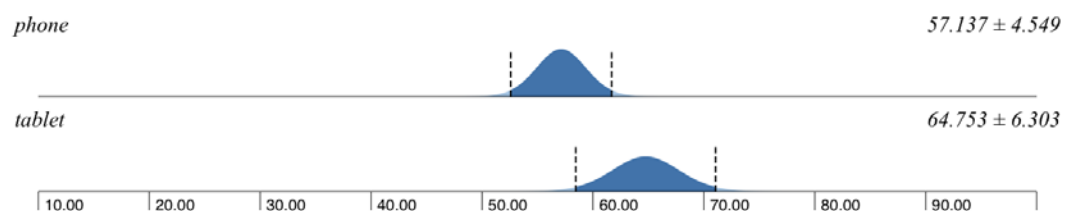


Figure 4-28: NASA TLX (mean) net task load score by device type

Group Statistics					
Device Type		N	Mean	Std. Deviation	Std. Error Mean
NASA TLX	Mobile Phone	36	57.13	13.45	2.24
Weighted Score	Tablet	24	64.75	14.92	3.04

Table 4-5 Group Statistics for NASA TLX Score by Device Type

Conclusion

The null hypothesis can be rejected. There was a statistically significant difference in NASA TLX task load score between the learners grouped by device type

The tablet-learners reported a higher overall task load than the smartphone-learners. The effect size was medium.

4.8.2. Comparing Test Scores between Learners Grouped by Device Type (H26)

Null Hypothesis:

H26: There will be no statistically significant difference in pre/post-test score reported between the learners grouped by device type (smartphone vs. tablet computer).

An independent-samples t-test was run to determine if there were differences in pre/post-test score between smartphone-learners and tablet-learners. There were no extreme outliers in the data, as assessed by inspection of a boxplot. Test scores for each group were normally distributed, as assessed by Shapiro-Wilk's test ($p > .05$), and there was homogeneity of variances, as assessed by Levene's test for equality of variances ($p=.422$). The test indicated no statistically significant difference between the smartphone-learners group ($M=24.25 \pm 4.27$) and the tablet-learners group ($M=27.37 \pm 5.88$), $t(58)=.0921$, $p=.371$.

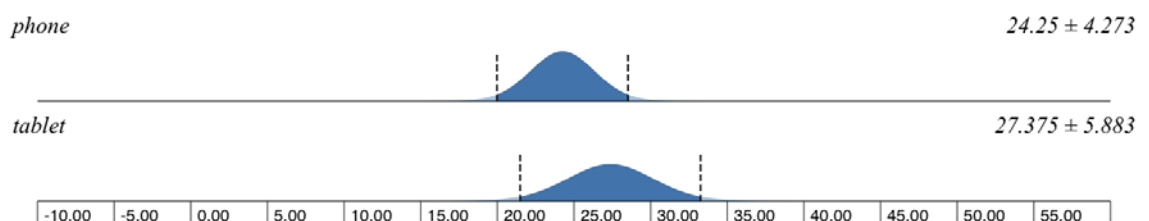


Figure 4-29 Pre/post-test scores by device type

Conclusion

The group means were not statistically significant different. The null hypothesis cannot be rejected.

4.8.3. Comparing Mental Demand Between Learners Grouped by Device Type (H27)

Null hypothesis:

H27: There will be no statistically significant difference in reported mental demand between the learners grouped by device type (smartphone vs. tablet computer).

Result

The data were not normally distributed in either group.

A Mann-Whitney U test was run to determine if there were differences in mental demand between smartphone-learners and tablet-learners. Distributions of the mental demand scores for mobile-learners and non-interactive learners were not similar, as assessed by visual inspection. Mental demand score for the tablet-learners group (mean rank=36.60) was statistically significantly higher than for the smartphone-learners (mean rank=26.43), $U=578.5$, $z=2.213$, $p=.027$, $r=.286$.

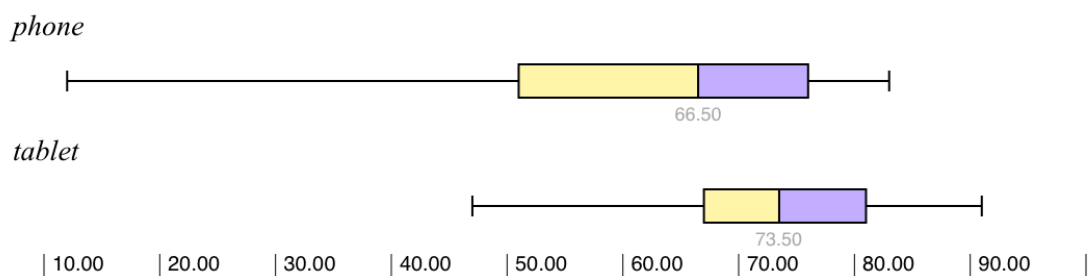


Figure 4-30: Boxplot of reported mental demand by device type (median values shown)

Conclusion

The null hypothesis can be rejected. There is a statistically significant difference in median-rank mental-demand score between the two groups. The tablet-learners reported a higher level of mental demand during the learning-activity than the smartphone-learners. The effect size was small.

4.8.4. Comparing Physical Demand Between Learners Grouped by Device Type (H28)

Null hypothesis:

H28: There will be no statistically significant difference in reported physical demand between the learners grouped by device type (smartphone vs. tablet computer).

Result:

The data were not normally distributed in either group.

A Mann-Whitney U test was run to determine if there were differences in physical demand between smartphone-learners and tablet-learners. Distributions of the physical demand scores for smartphone-learners and tablet-learners were not similar, as assessed by visual inspection. Physical demand score for the tablet-learners group (mean rank=36.50) was statistically significantly higher than for the smartphone-learners (mean rank=26.50), $U=576.0$, $z=2.176$, $p=.030$, $r=.281$.

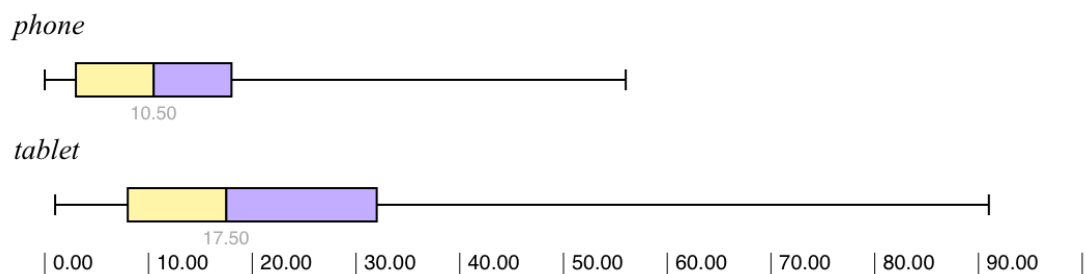


Figure 4-31: Boxplot of reported physical demand by device type (median values shown)

Conclusion

The null hypothesis can be rejected. There is a statistically significant difference in median-rank physical-demand score between the two groups. The tablet-learners reported a higher level of physical demand during the learning-activity than the smartphone-learners. The effect size was small.

4.8.5. Comparing Temporal Demand Between Learners Grouped by Device Type (H29)

Null hypothesis:

H29: There will be no statistically significant difference in reported temporal demand between the learners grouped by device type (smartphone vs. tablet computer).

Result:

An independent-samples t-test was run to determine if there were differences in temporal demand between smartphone-learners and tablet-learners. There were no extreme outliers in the data, as assessed by inspection of a boxplot. Temporal demand scores for each group were normally distributed, as assessed by Shapiro-Wilk's test ($p > .05$). The mean difference between smartphone-learners (61.31 ± 6.61) and tablet-learners (63.87 ± 7.36) was not statistically significant $p = .605$.

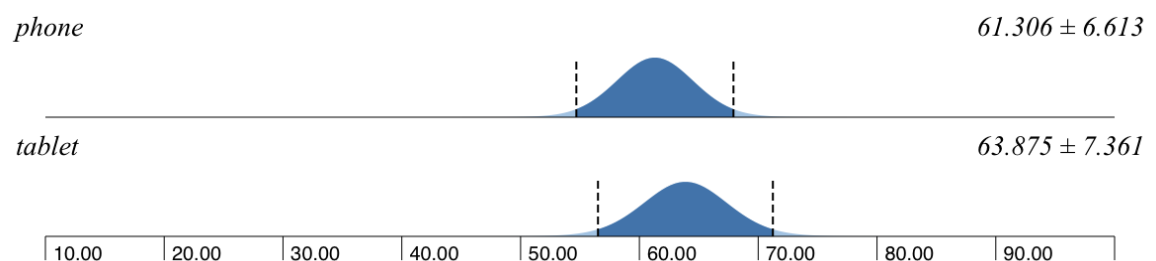


Figure 4-32: Temporal Demand by Device Type

Conclusion:

The group means were not statistically significant different. The null hypothesis cannot be rejected.

4.8.6. Comparing Performance Between Learners Grouped by Device Type (H30)

Null hypothesis:

H30: There will be no statistically significant difference in reported performance between the learners grouped by device type (smartphone vs. tablet computer).

Result:

An independent-samples t-test was run to determine if there were differences in performance between smartphone-learners and tablet-learners. There were no extreme outliers in the data, as assessed by inspection of a boxplot. Performance scores for each group were normally distributed, as assessed by Shapiro-Wilk's test ($p > .05$). The mean difference in performance between smartphone-learners (45.14 ± 5.97) and tablet-learners (51.37 ± 10.16) was not statistically significant $p=.251$.

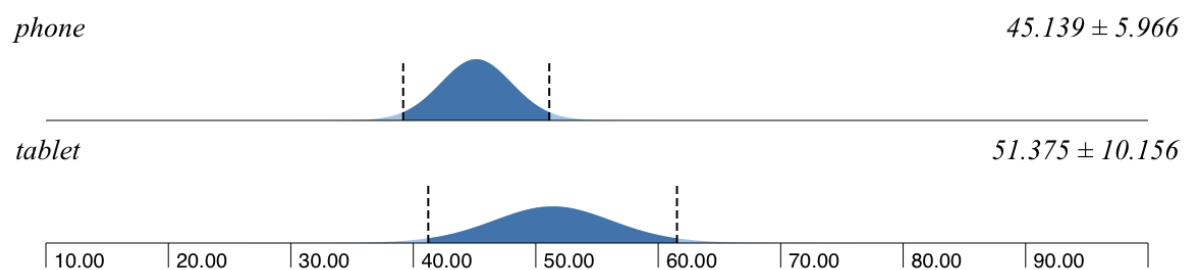


Figure 4-33: Performance by device type

Conclusion:

The group means were not statistically significant different. The null hypothesis cannot be rejected.

4.8.7. Comparing Effort Between Learners Grouped by Device Type (H31)

Null hypothesis:

H31: There will be no statistically significant difference in reported effort between the learners grouped by device type (smartphone vs. tablet computer).

Result

An independent-samples t-test was run to determine if there were differences in effort between smartphone-learners and tablet-learners. There were no extreme outliers in the data, as assessed by inspection of a boxplot. Effort scores for each group were normally distributed, as assessed by Shapiro-Wilk's test ($p > .05$), and there was homogeneity of variances, as assessed by Levene's test for equality of variances ($p=.964$). The reported effort was greater for the tablet-learners group (71.00 ± 8.12) than for the smartphone-learners group (60.03 ± 6.57), a statistically significant difference of -10.97 (95% CI, -0.76 to -21.17), $t(60)=2.153$, $p=.036$, $d=.565$.

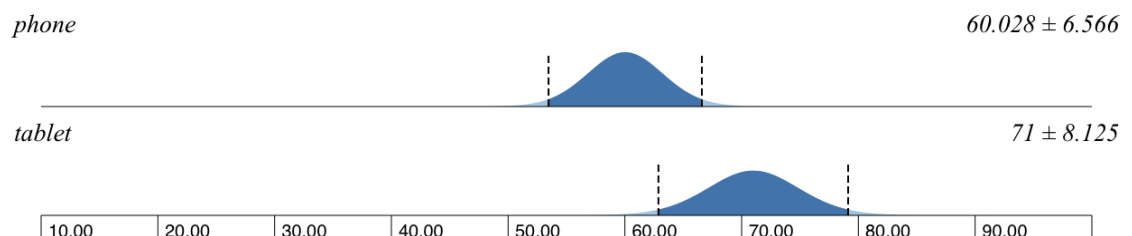


Figure 4-34: Effort by device type

Group		N	Mean	Std. Deviation	Std. Error Mean
Reported	Mobile Phones	36	60.02	19.40	3.23
Effort	Tablets	24	71.00	19.24	3.92

Table 4-6: Group Statistics for effort by Device Type

Conclusion

The null hypothesis can be rejected. The group means were statistically significantly different. The tablet-learners reported a higher level of effort than the smartphone-learners. The effect size was medium.

4.8.8. Comparing Frustration Between Learners Grouped by Device Type (H32)

Null hypothesis:

H32: There will be no statistically significant difference in reported frustration between the learners grouped by device type (smartphone vs. tablet computer).

Result

An independent-samples t-test was run to determine if there were differences in frustration between smartphone-learners and tablet-learners. There were no extreme outliers in the data, as assessed by inspection of a boxplot. Frustration scores for each group were normally distributed, as assessed by Shapiro-Wilk's test ($p > .05$), and there was homogeneity of variances, as assessed by Levene's test for equality of variances ($p=.411$). The mean difference in frustration between smartphone-learners (41.306 ± 7.816) and tablet-learners (50.625 ± 11.21) was not statistically significant $p=.155$.

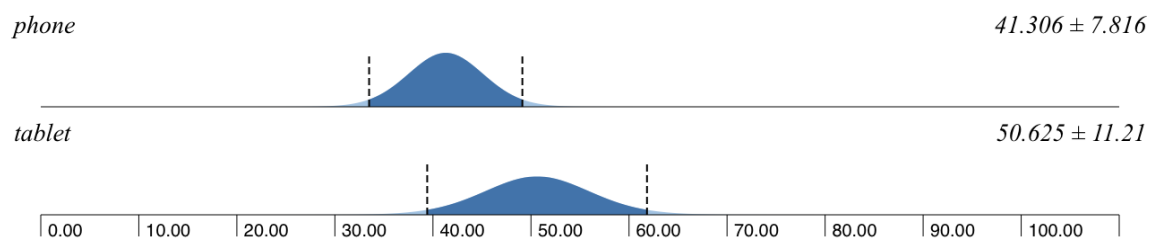


Figure 4-35: Frustration by device type

Conclusion

The group means were not statistically significant different. The null hypothesis cannot be rejected.

4.8.9. Summary of Results for Comparisons between Device Type (Phone vs. Tablet)

Greyed out data refers to results obtained from tests that were not optimal for the data under analysis but were undertaken for comparison purposes. All scores are out of 100. Text highlighted in blue refers to data from non-parametric tests and indicates the median score rather than the mean. Asterisks indicate statistically significant results ($p < .05$).

Dependent Variables	Smartphone Learners ($n=36$) Score <i>M</i> (SD)/ <i>Mdn</i> (IQR)	Tablet Learners ($n=24$) Score <i>M</i> (SD)/ <i>Mdn</i> (IQR)	Mann-Whitney p-value	Mann-Whitney CI(%)	t-test p-value	t-test CI (%)	Sig.	Effect Size
(H26)Pre-Post Test score (%)	24.25 (12.63)	27.37 (14.92)	0.283	<80	0.371	<80	No	
(H25) Net Task load	57.14 (13.45)	64.75 (14.92)	0.096	92	0.044	95	Yes	$d=.566$
(H27)Mental Demand	66.50 (24)	73.50 (14)*	0.027	97	0.010	98	Yes	$r=.286$
(H28) Physical Demand	10.50 (15)	17.50 (25)*	0.030	97	0.019	98	Yes	$r=.281$
(H29) Temporal Demand	61.31 (19.54)	63.87 (17.43)	0.544	<80	0.605	<80	No	
(H30) Performance	45.14 (17.63)	51.37 (24.05)	0.422	<80	0.251	<80	No	
(H31) Effort	60.03 (19.40)	71.00 (19.24)	0.017	98	0.036	96	Yes	$d=.565$
(H32) Frustration	41.30 (23.10)	50.62 (26.55)	0.193	80	0.155	84	No	

Table 4-7: Summary of all comparisons relating to device type

The table shows that the tablet PC learners reported a statistically significantly higher level of net task load, mental demand, physical demand and effort than the smartphone-learners.

4.9. RESULTS RELATING TO SCREEN SIZE

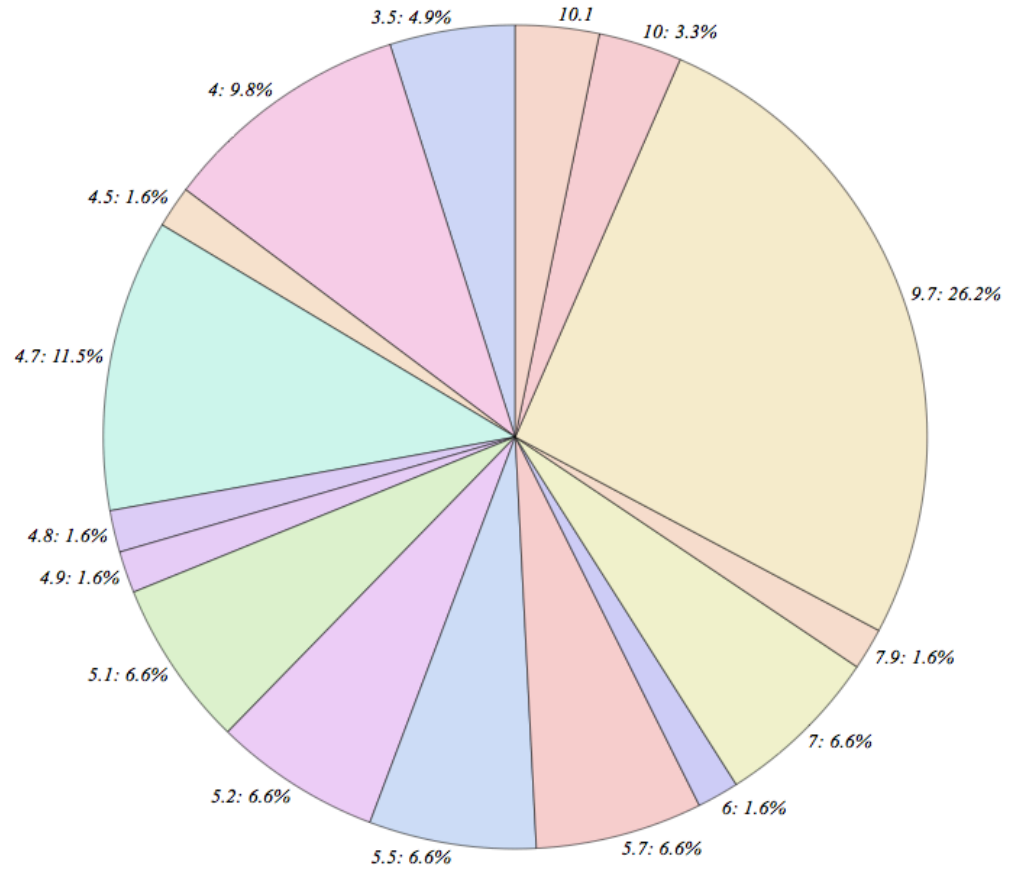


Figure 4-36: Pie chart showing distribution of mobile device screen sizes in inches

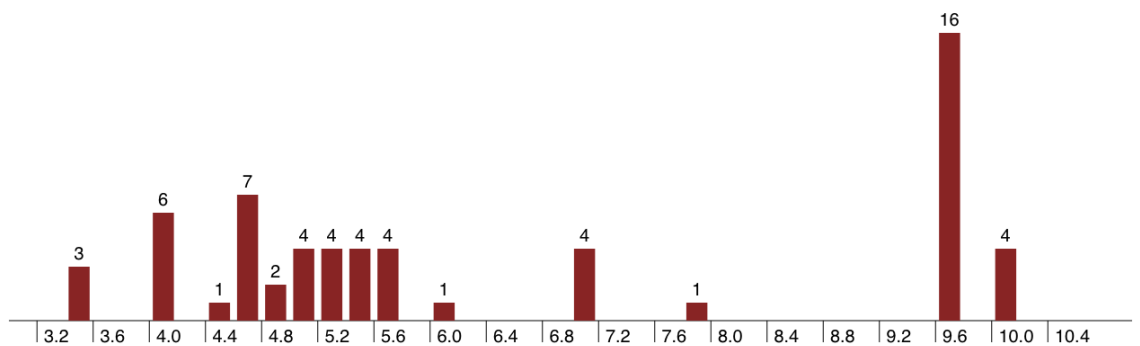


Figure 4-37: Histogram showing distribution of mobile device screen sizes in inches

To analyse the data it was necessary to group participants into meaningful categories in terms of the values that represent the screen-size for typical types of device. These groups represent the typical screen sizes of:

Small - Early smartphones models such as Apple iPhone (3.5" - 4.5")

Medium - Recent smartphone models Apple, Google, Sony and Samsung (4.7" – 5.4")

Large - "Phablet" devices - large smartphones and small tablets (iPad Mini) (5.5" – 9.6")

Very Large - Tablet computers such as the iPad (9.7" - 10.1")

This strategy allowed ANOVA or Kruskal-Wallis tests to be run on all of the independent variables. The rationale for grouping the participants in this way is because a simple correlation between screen size and the dimensions of the NASA TLX does not always reveal important trends. This can be seen on page 208 where the ANOVA reveals that there is a statistically significant difference in task load between large smartphones and tablets. A correlation between screen size and task load does not show any statistical significance across the whole range of sizes.

4.9.1. Comparing Net Task Load Between Learners Grouped by Screen Size (H33)

Null Hypothesis:

H33: There will be no statistically significant difference in NASA TLX task load score between the learners grouped by screen size of device used.

Result:

A one-way ANOVA was run to determine if there were differences in NASA TLX task load between participants grouped by device screen-size. There were no extreme outliers in the data, as assessed by inspection of a boxplot. Task load scores for each group were normally distributed, as assessed by Shapiro-Wilk's test ($p > .05$).

NASA TLX task load score was statistically significantly different between different screen sizes, $F(3, 56)=2.594$, $p=.049$, $\omega^2=.121$.

NASA TLX task load score increased from large screen size (5.4"-9.7") (52.96 ± 7.40), to medium screen sizes (4.7"-5.4") (57.81 ± 8.00), to small screen sizes (3.5"- 4.6") (55.65 ± 12.12). However, the NASA TLX task load score for very large screen size tablets (9.7"- 10.1") (66.1 ± 7.40) was greater than that for the large screen smartphones.

Tukey post hoc analysis revealed that the increase from large to very large screen sizes (-13.13, 95% CI (-26.23 to -0.04) was statistically significant with a large effect size ($p=.049$, $d=1.131$), but no other group differences were statistically significant.

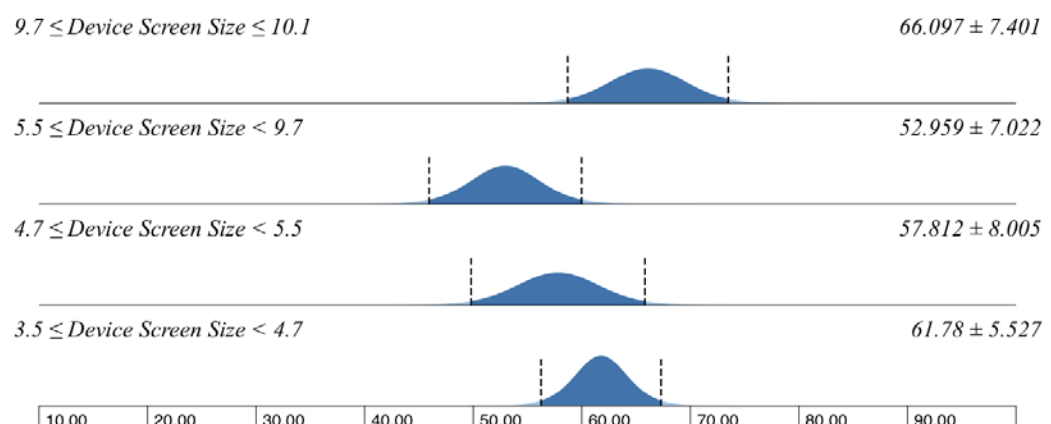


Figure 4-38 NASA TLX (mean) score by Device Screen Size

Conclusion:

The null hypothesis can be rejected. The group means were statistically significantly different. The tablet-learners reported a higher level of task load during the learning task than the smartphone-learners using devices with large screens. The effect size was large.

4.9.2. Comparing Test-Scores Between Learners Grouped by Screen Size (H34)

Null Hypothesis:

H34: There will be no statistically significant difference in pre/post-test scores between the learners grouped by screen size of device used.

Result:

The data followed a normal distribution.

There was no statistically significant difference between groups as determined by a one-way ANOVA ($p=.748$).

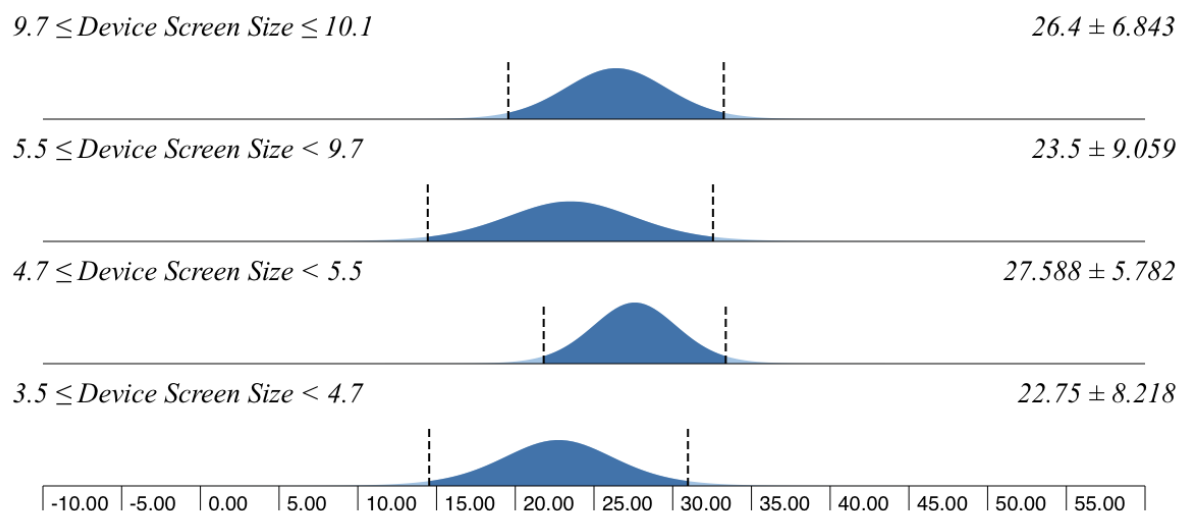


Figure 4-39: Pre/post test score by screen size

Conclusion

The group means were not statistically significant different. The null hypothesis cannot be rejected.

4.9.3. Comparing Mental Demand Between Learners Grouped by Screen Size (H35)

Null Hypothesis:

H35: There will be no statistically significant difference in reported mental demand between the learners grouped by screen size of device used.

Result:

The data did not follow normal distribution as assessed by Shapiro-Wilk's test ($p > .05$). There was no statistically significant difference between groups as determined by a Kruskal-Wallis test ($p=.062$).

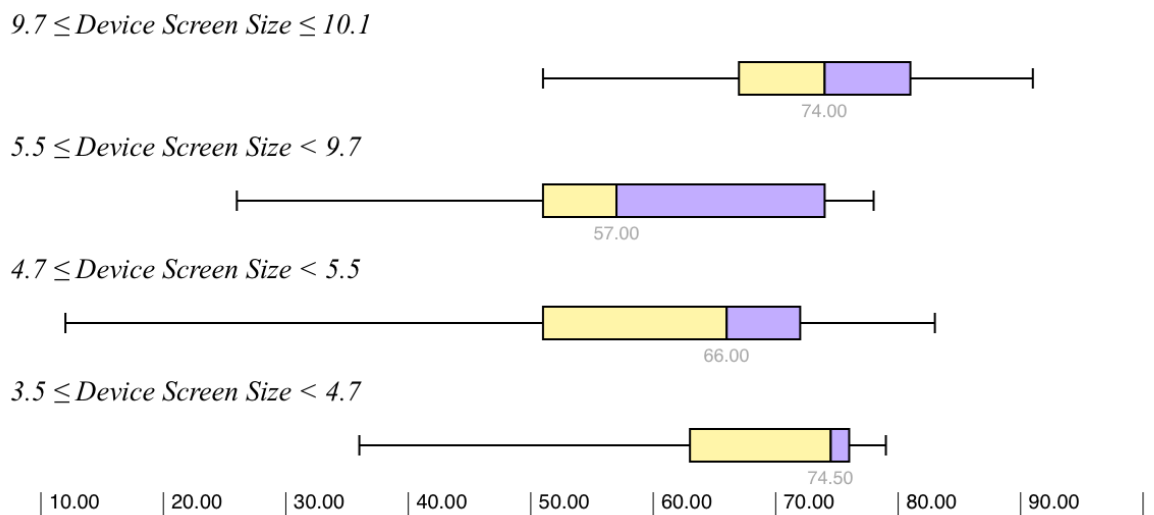


Figure 4-40: Boxplot of mental demand by screen-size (median values shown)

Conclusion

There was no statistically significant difference in median mental-demand score between groups. The null hypothesis cannot be rejected.

4.9.4. Comparing Physical Demand Between Learners Grouped by Screen Size (H36)

Null Hypothesis:

H36: There will be no statistically significant difference in reported physical demand between the learners grouped by screen size of device used.

Result:

The data did not follow normal distribution as assessed by Shapiro-Wilk's test ($p > .05$). There was no statistically significant difference between groups as determined by a Kruskal-Wallis test ($p=.203$) or one-way ANOVA ($p=.118$).

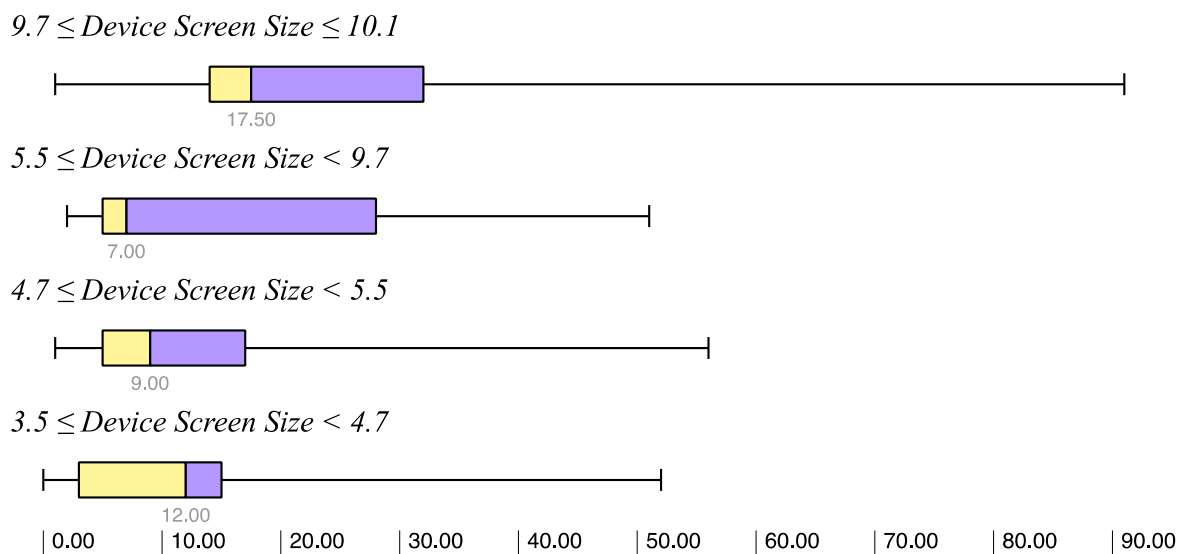


Figure 4-41: Boxplot of physical demand by screen size (median values shown)

Conclusion

There was no statistically significant difference in median physical-demand score between groups. The null hypothesis cannot be rejected.

4.9.5. Comparing Temporal Demand Between Learners Grouped by Screen Size (H37)

Null Hypothesis:

H37: There will be no statistically significant difference in reported temporal demand between the learners grouped by screen size of device used.

Result:

The data did not follow normal distribution as assessed by Shapiro-Wilk's test ($p > .05$). There was no statistically significant difference between groups as determined by a Kruskal-Wallis test ($p = .501$) or one-way ANOVA ($p = .480$).

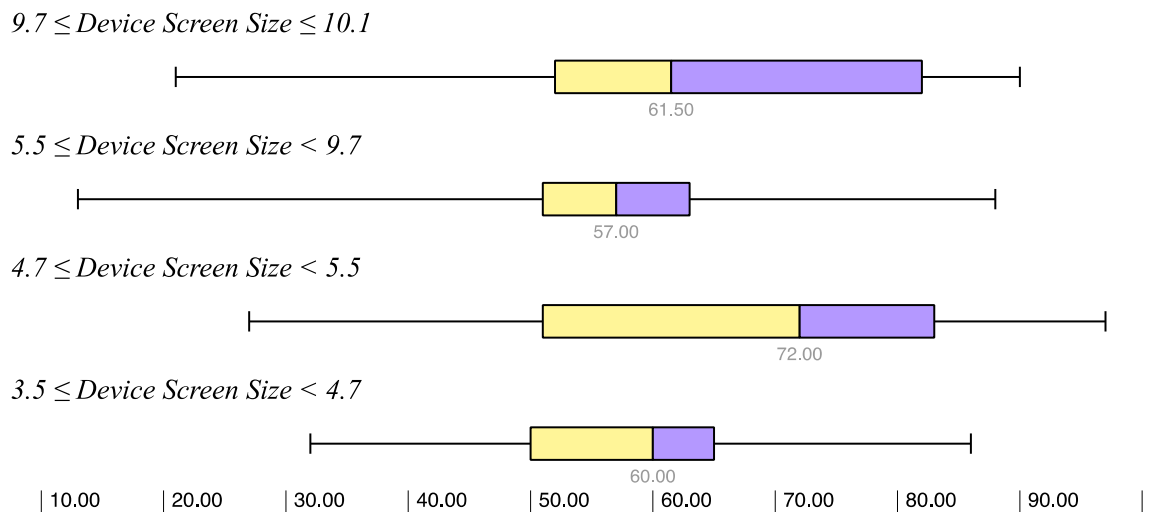


Figure 4-42: Boxplot of temporal demand by screen size (median values shown)

Conclusion:

There was no statistically significant difference in median temporal-demand score between groups. The null hypothesis cannot be rejected.

4.9.6. Comparing Performance Between Learners Grouped by Screen Size (H38)

Null Hypothesis:

H38: There will be no statistically significant difference in reported performance between the learners grouped by screen size of device used.

Result:

A one-way ANOVA was run to determine if there were differences in performance between participants grouped by device screen-size. There were no extreme outliers in the data, as assessed by inspection of a boxplot. Performance scores for each group were normally distributed, as assessed by Shapiro-Wilk's test ($p > .05$).

Performance score was not statistically significantly different between different screen sizes, $F(3, 56) = .968, p = .415$.

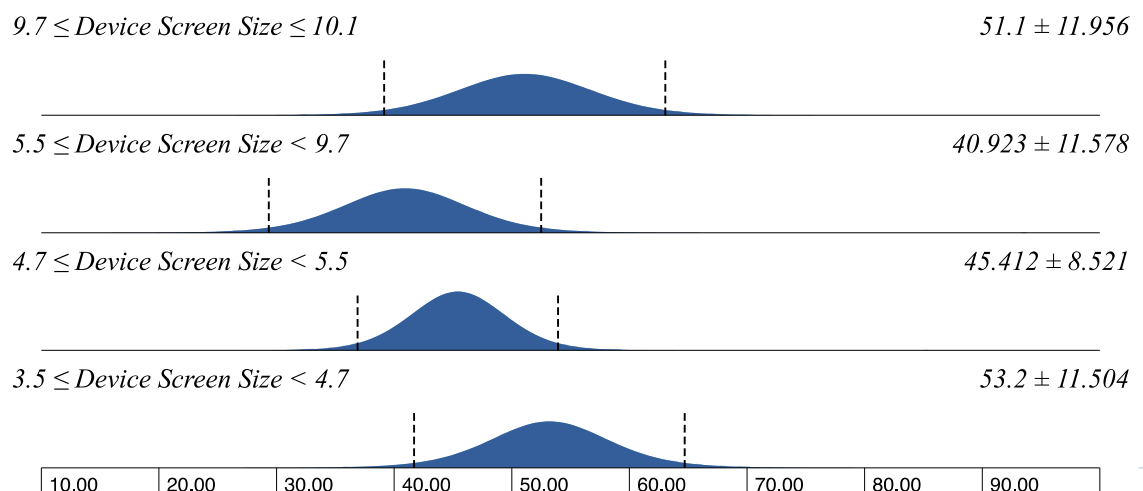


Figure 4-43: Performance by device screen size

Conclusion:

The group means were not statistically significant different. The null hypothesis cannot be rejected.

4.9.7. Comparing Effort Between Learners Grouped by Screen Size (H39)

Null Hypothesis:

H39: There will be no statistically significant difference in reported effort between the learners grouped by screen size of device used.

Result:

The data did not follow normal distribution as assessed by Shapiro-Wilk's test ($p > .05$). A Kruskal-Wallis H test was run to determine if there were differences in effort score between four groups of participants with different device screen-sizes: the "small (1)", "medium (2)", "large (3)" and "very large(4)" groups. Distributions of effort scores were not similar for all groups, as assessed by visual inspection of a boxplot. The mean ranks of effort scores were statistically significantly different between groups, $\chi^2(3)=8.459$, $p=.037$. Pairwise comparisons were performed using Dunn's (1964) procedure with a Bonferroni correction for multiple comparisons. Adjusted p-values are presented. This post hoc analysis revealed statistically significant differences in effort mean-rank scores between the medium (22.59) and very large (38.52) ($p=.034$, $\eta^2=.143$) groups, but not between any other group combination.

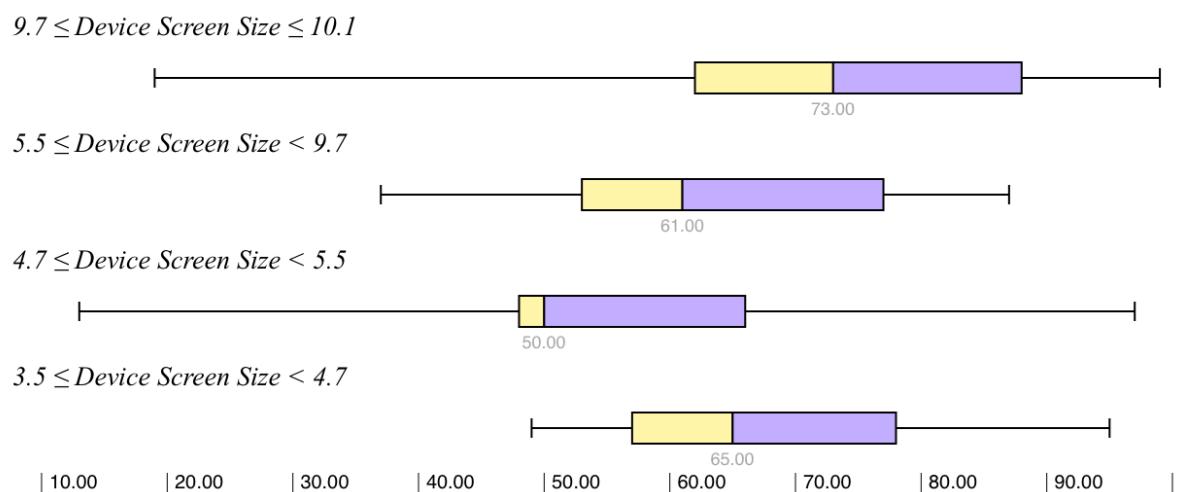
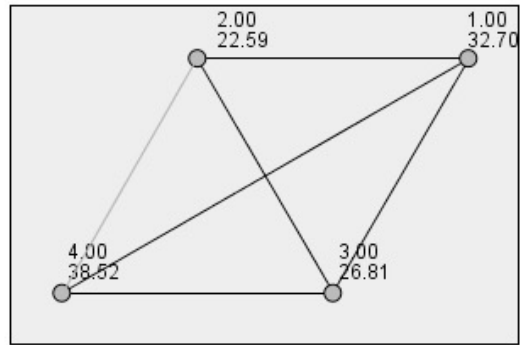


Figure 4-44: Boxplot of reported effort by device screen size (median scores shown)



Each node shows the sample average rank of screen_size_code.

Figure 4-45: Pairwise comparisons of effort by device screen size

Sample1-Sample2	Test Statistic	Std. Error	Std. Test Statistic	Sig.	Adj.Sig.
2.00-4.00	-15.937	5.759	-2.767	.006	.034
3.00-4.00	-11.717	6.219	-1.884	.060	.357
2.00-1.00	10.112	6.957	1.453	.146	.877
1.00-4.00	-5.825	6.761	-.862	.389	1.000
2.00-3.00	-4.219	6.432	-.656	.512	1.000
3.00-1.00	5.892	7.343	.802	.422	1.000

Table 4-8: Statistically significant pairwise comparisons of effort by device screen size

Conclusion

The null hypothesis can be rejected. There is a statistically significant difference in median-rank effort score between the two groups. The tablet-learners (very large screen size) reported a higher level of effort during the learning task than the smartphone-learners using a medium screen size. The effect size was medium.

4.9.8. Comparing Frustration Between Learners Grouped by Screen Size (H40)

Null Hypothesis:

H40: There will be no statistically significant difference in reported frustration between the learners grouped by screen size.

Result:

A one-way ANOVA was conducted to determine if the level of frustration was different for groups using devices of different screen size.

There were no outliers, as assessed by boxplot; data were normally distributed for each group, as assessed by Shapiro-Wilk test ($p > .05$); and there was homogeneity of variances, as assessed by Levene's test of homogeneity of variances ($p=.465$).

The frustration score was statistically significantly different between different screen sizes used, $F(3, 56)=3.770$, $p=.016$, $\omega^2=.121$. Frustration score increased from small screen size (5.5"-9.6") (29.76 \pm 13.2), to medium screen sizes (9.7"-10.1") (55.65 \pm 12.12), and increased from large screen size (5.5"- 9.6") (29.76 \pm 13.2), to very large (9.7"-10.1") (55.65 \pm 12.12). However, the frustration score for small screen sizes was greater than that for very large screen sizes.

Tukey post hoc analysis revealed that the increase from large to very large screen sizes (25.88, 95% CI (4.0229 to 47.7389) was statistically significant ($p=.014$, $d=1.186$), but no other group differences were statistically significant.

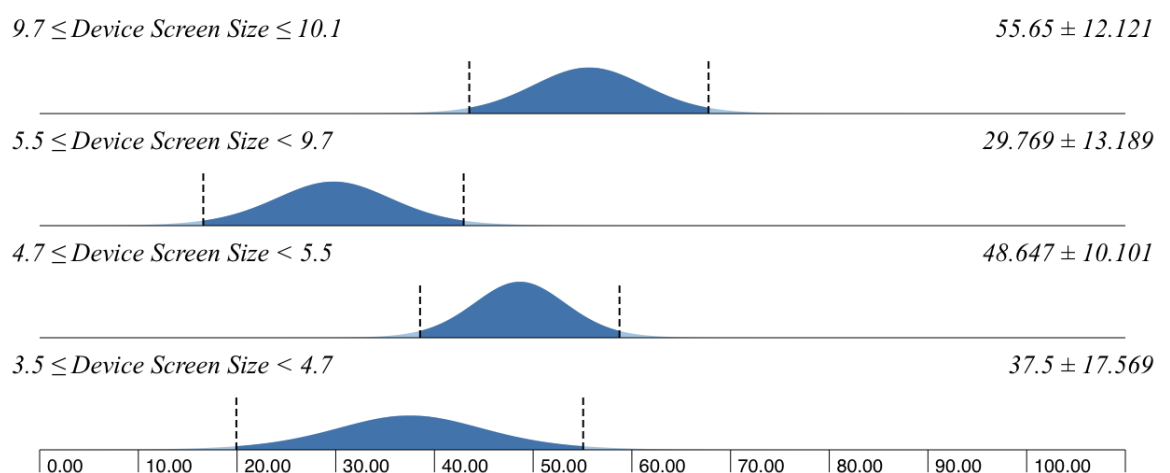


Figure 4-46: Reported frustration by device screen-size

Multiple Comparisons						
Dependent Variable: Frustration						
Tukey HSD						
(I) screen-size	(J) screen-size	Mean Difference (I-J) *Significant at the 0.5 level	Std. Error	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
3.5" – 4.5"	4.7"- 5.5"	-11.14	9.23	.625	-35.59	13.30
	5.5" – 7.0"	7.73	9.74	.857	-18.07	33.53
	9.7" – 10.1"	-18.15	8.97	.192	-41.91	5.61
4.7" – 5.5"	5.5" – 7.0"	18.87	8.53	.133	-3.72	41.48
	9.7" – 10.1"	-7.00	7.64	.796	-27.24	13.23
5.5" – 7.0"	9.7" – 10.1"	-25.88*	8.25	.014	-47.73	-4.02

Table 4-9: Multiple comparisons of frustration by screen-size (Tukey HSD)

Conclusion:

The null hypothesis can be rejected. The mean frustration scores were statistically significantly different between groups. The tablet-learners reported a higher level of frustration during the learning task than the smartphone-learners using devices with large screens. The effect size was large.

4.9.9. Summary of Results for Comparisons between Learners Grouped by Screen-Size

Text highlighted in blue refers to data from non-parametric tests and indicates the median score rather than the mean.

Asterisks indicate statistically significant results ($p < .05$).

Dependent Variables	Small (n=10) score <i>M (SD)/Mdn (IQR)</i>	Medium (n=17) score <i>M (SD)/Mdn (IQR)</i>	Large (n=13) score <i>M (SD)/Mdn (IQR)</i>	V large (n=20) score <i>M (SD)/Mdn (IQR)</i>	p-value ANOVA Kruskal-Wallis	CI (%)	Sig?	Effect size
(H34) Pre-Post Test score (%)	22.75 (11.49)	27.59 (11.24)	23.50 (14.99)	26.40 (14.62)	0.748	<80	No	
(H33) Net Task load	61.78 (7.73)	57.81 (15.57)	52.96 (11.62)*	66.10 (15.81)*	0.049	95	Yes	$d=1.131$
(H35) Mental Demand	74.50 (14)	66.00 (21)	57.00 (23)	74.00 (14)	0.062	93	No	
(H36) Physical Demand	12.00 (19)	9.00 (12)	7.00 (22)	17.50 (17)	0.203	<80	No	
(H37) Temporal Demand	60.00 (27)	72.00 (32)	57.00 (12)	61.50 (28)	0.501	<80	No	
(H38) Performance	53.20 (16.08)	45.41 (16.57)	40.92 (19.16)	51.10 (25.55)	0.415	<80	No	
(H39) Effort	65.00 (22)	50.00 (18)*	61.00 (24)	73.00 (25)*	0.034	95	Yes	$\eta^2=.143$
(H40) Frustration	37.50 (24.56)	48.65 (19.65)	29.77 (21.83)*	55.65 (25.90)*	0.014	98	Yes	$d=1.186$

Table 4-10: Summary of all results for comparisons between learners grouped by device screen size

The table shows that learners having devices with very large screens (tablets having screen sizes between 9.7 and 10.1 inches) reported a statistically significantly higher level of net task load and frustration than learners having large screens (phablets between 5.5 and 9.6 inches). They also reported a higher level of effort than learners using medium sized screens (smartphones having a screen size of between 4.7 and 5.4 inches).

4.10. RESULTS RELATING TO SPATIAL RESOLUTION OF DEVICE DISPLAY

For this part of the study, learners were divided into two groups. Group 1 ($n=36$), (retina display), were participants using devices having a display resolution of greater than 290 PPI. Group 2 ($n=24$), (non-retina-display), were participants using devices having a display resolution of fewer than 290 PPI. Tests were run for all of the dependent variables.

4.10.1. Comparing Net Task load Between Learners Grouped by Screen Resolution (H41)

Null hypothesis:

H41: There will be no statistically significant difference in NASA TLX task load score between learners grouped by spatial resolution of device screen.

Result:

An independent-samples t-test was run to determine if there were differences in NASA TLX task load score between learners using retina display devices and non-retina-display devices. There were no extreme outliers in the data, as assessed by inspection of a boxplot. Task load scores for each group were normally distributed, as assessed by Shapiro-Wilk's test ($p > .05$), and there was homogeneity of variances, as assessed by Levene's test for equality of variances ($p=.347$). The test indicated no statistically significant difference between the retina display group ($M=58.63 \pm 4.66$) and the non-retina-display group ($M=62.50 \pm 6.48$), $t(60)=1.58$, $p=.312$.

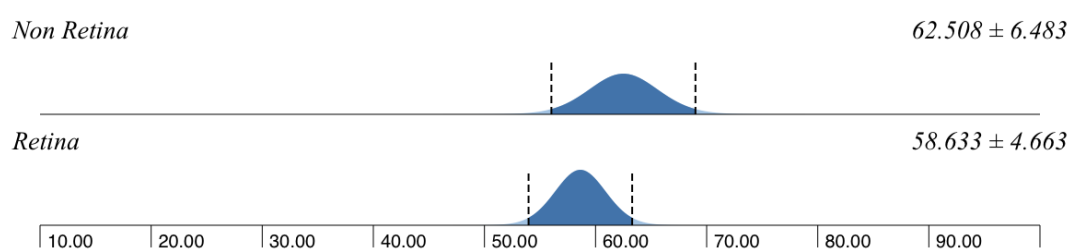


Figure 4-47: NASA TLX (mean) score by spatial resolution of device display

Conclusion:

The group means were not statistically significant different. The null hypothesis cannot be rejected.

A Pearson's product-moment correlation was run to assess the relationship between NASA TLX net task load and screen resolution in pixels-per-inch.

Result:

There was a weak negative correlation between NASA TLX net task load and screen resolution in pixels-per-inch, $z(60)=-.266$, $p=.042$.

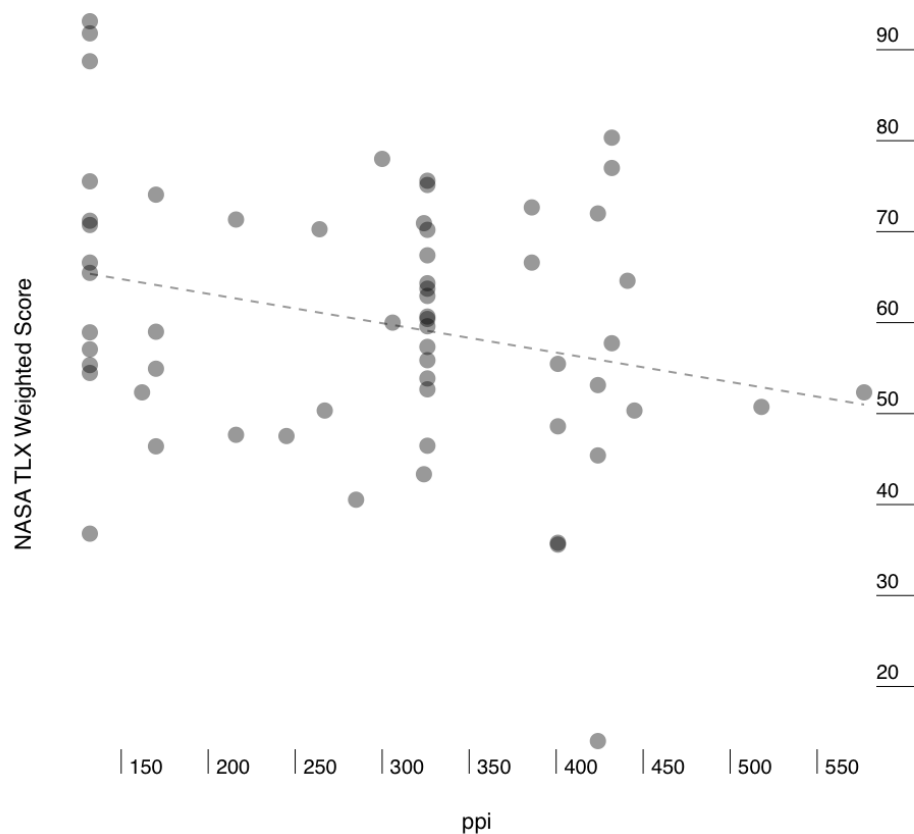


Figure 4-48: Scatterplot showing weak negative correlation between NASA TLX net task load and device screen resolution

4.10.2. Comparing Test-Scores Between Learners Grouped by Screen Resolution (H42)

Null Hypothesis:

H42: There will be no statistically significant difference in pre/post-test scores between the learners grouped by spatial resolution of device display.

Result:

An independent-samples t-test was run to determine if there were differences in pre/post-test scores between learners using retina display devices and non-retina-display devices. There were no extreme outliers in the data, as assessed by inspection of a boxplot. Test scores for each group were normally distributed, as assessed by Shapiro-Wilk's test ($p > .05$), and there was homogeneity of variances, as assessed by Levene's test for equality of variances ($p=.073$). The test indicated no statistically significant difference between the retina display group ($M=25.22 \pm 3.47$) and the non-retina-display group ($M=25.92 \pm 5.42$), $t(60)=.199$, $p=.843$.

Conclusion:

The group means were not statistically significant different. The null hypothesis cannot be rejected.

A Pearson's product-moment correlation was performed to assess the relationship between pre/post test score and screen resolution in pixels-per-inch.

Result:

There was no statistically significant correlation found.

4.10.3. Comparing Mental Demand Between Learners Grouped by Screen Resolution (H43)

Null Hypothesis:

H43: There will be no statistically significant difference in reported mental demand between the learners grouped by spatial resolution of device display.

Result:

A Shapiro-Wilk Test indicated that the data were not normally distributed in the retina-display learners group ($p=.001$).

A Mann-Whitney U test was run to determine if there were differences in mental demand between learners using retina-display devices and learners using non-retina-display devices learners. Distributions of the mental demand scores for mobile-learners and non-interactive learners were not similar, as assessed by visual inspection. Median mental demand score for learners using retina-display devices (mean rank=29.28) and learners using non-retina-display devices (mean rank=32.33) was not statistically significantly different, $U=388$, $z=.665$, $p=.506$

Conclusion

There was no statistically significant difference between mean rank mental demand scores between the groups The null hypothesis cannot be rejected.

A Pearson's product-moment correlation was performed to assess the relationship between mental demand and screen resolution in pixels-per-inch.

Result

There was a weak negative correlation between mental demand and screen resolution in pixels-per-inch, $z(60)=-.329$, $p=.010$.

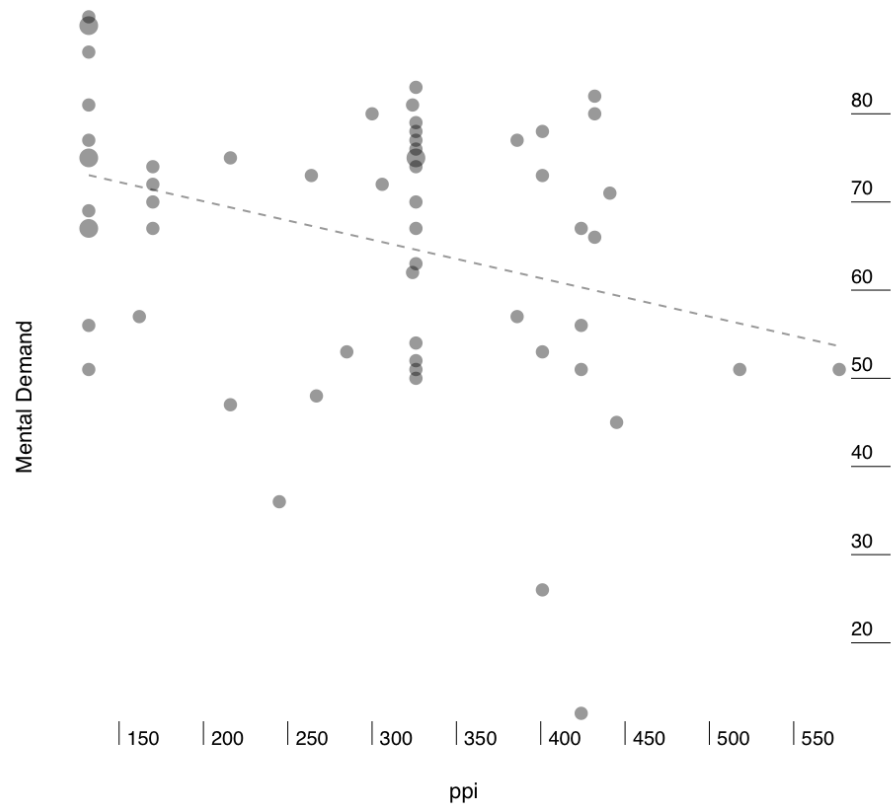


Figure 4-49: Scatterplot showing a weak negative correlation between mental demand and device screen resolution

4.10.4. Comparing Physical Demand Between Learners Grouped by Screen Resolution (H44)

Null Hypothesis:

H44: There will be no statistically significant difference in reported physical demand between the learners grouped by spatial resolution of device display.

Result:

The Shapiro-Wilk Test indicated that the physical demand data were not normally distributed in either group (retina $p=.001$, non-retina $p=.002$) A Mann-Whitney U test was run to determine if there were differences in physical demand between learners using retina-display devices and learners using non-retina-display devices. Distributions of physical demand for mobile-learners and non-interactive learners were similar, as assessed by visual inspection. Median physical demand score was statistically significantly higher in learners using non-retina-display devices (17.50) than in learners using retina-display devices (10.50), $U=294.50$, $z=2.077$, $p=.038$. $r=.268$

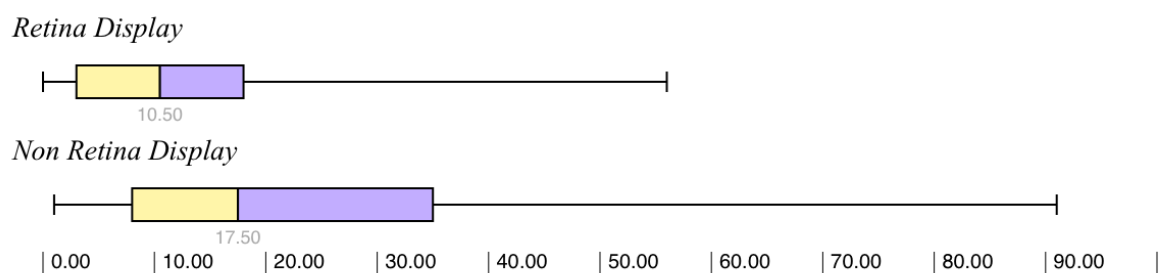


Figure 4-50: Boxplot of Physical Demand by Display Resolution (median values shown)

Conclusion

The null hypothesis can be rejected. There is a statistically significant difference in median physical-demand score between the two groups. Mobile-learners using non-“retina-display” devices having a screen resolution of fewer than 290 PPI reported a higher level

of physical demand during the learning task than learners having a “retina-display” screen resolution of greater than 290 PPI. The effect size was small.

A Pearson's product-moment correlation was performed to assess the relationship between physical demand and screen resolution in pixels-per-inch.

Result

There was a weak negative correlation between physical demand and screen resolution in pixels-per-inch, $z(60)=-.298, p=.022$.

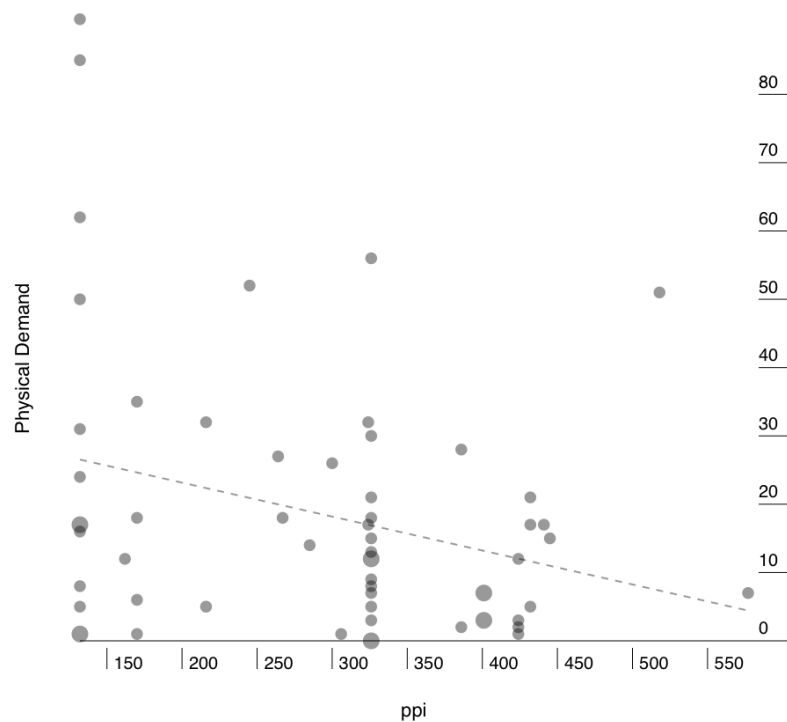


Figure 4-51: Scatterplot showing weak negative correlation between physical demand and device screen resolution

A Pearson's product-moment correlation was performed to assess the relationship between physical demand and temporal demand in both the non-retina-display group and retina-display group of learners. This test was performed to support the contention

that content zooming and swiping might have occupied task-time and increased temporal load in the non-retina-display group.

Result:

There was a moderate positive correlation between physical demand and temporal demand in the non-retina-display group, $z(60)=-.632$, $p=.004$.

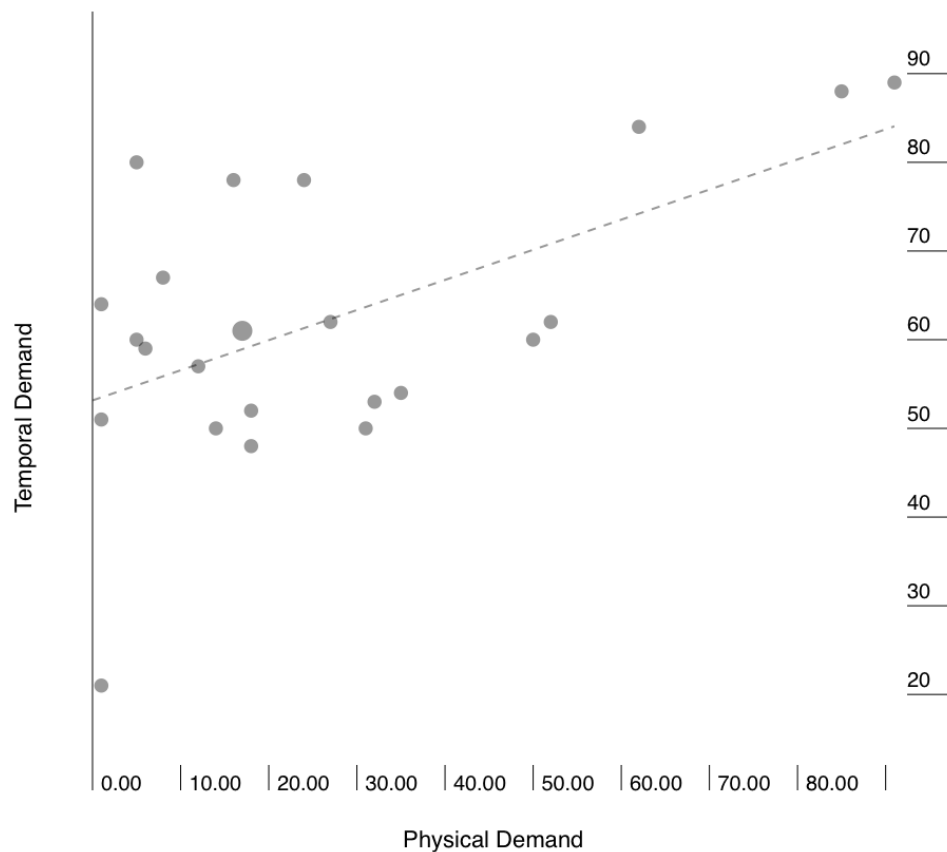


Figure 4-52: Scatterplot showing moderate positive correlation between physical demand and temporal demand in the non-retina-display group

There was no statistically significant correlation between physical demand and temporal demand in the retina-display group.

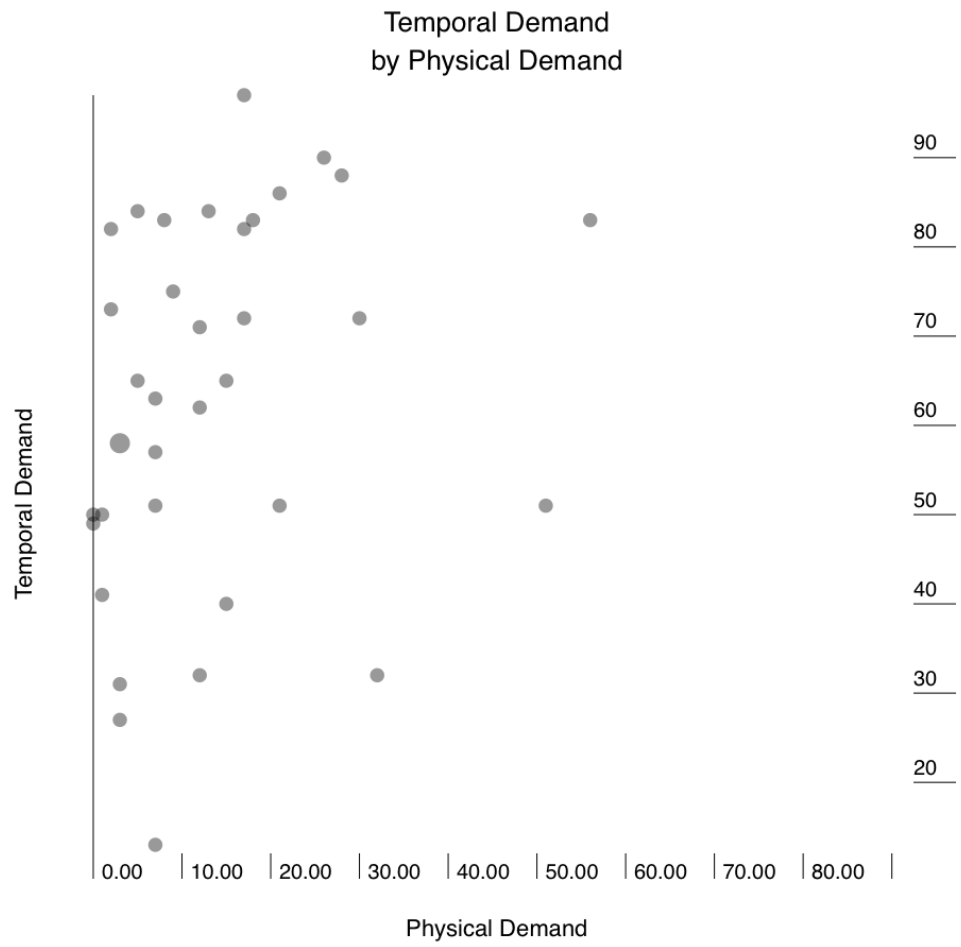


Figure 4-53: Scatterplot showing no significant correlation between physical demand and temporal demand in the retina-display group

4.10.5. Comparing Temporal Demand Between Learners Grouped by Screen Resolution (H45)

Null Hypothesis:

H45: There will be no statistically significant difference in reported temporal demand between the learners grouped by spatial resolution of device display.

Result:

An independent-samples t-test was run to determine if there were differences in temporal demand between learners using retina display devices and non-retina-display devices. There were no extreme outliers in the data, as assessed by inspection of a boxplot. Temporal demand scores for each group were normally distributed, as assessed by Shapiro-Wilk's test ($p > .05$), and there was homogeneity of variances, as assessed by Levene's test for equality of variances ($p=.34$). The test indicated no statistically significant difference between the retina display group (62.55 ± 7.03) and the non-retina-display group ($M=62.04 \pm 6.43$), $t(60)=.098$, $p=.922$.

Conclusion:

There was no statistically significant difference between means. The null hypothesis cannot be rejected.

A Pearson's product-moment correlation was performed to assess the relationship between temporal demand and screen resolution in pixels-per-inch.

Result:

There was no statistically significant correlation found.

4.10.6. Comparing Performance Between Learners Grouped by Screen Resolution (H46)

Null Hypothesis:

H46: There will be no statistically significant difference in reported performance between the learners grouped by spatial resolution of device display.

Result:

An independent-samples t-test was run to determine if there were differences in performance between learners using retina display devices and non-retina-display devices. There were no extreme outliers in the data, as assessed by inspection of a boxplot. Performance scores for each group were normally distributed, as assessed by Shapiro-Wilk's test ($p > .05$), and there was homogeneity of variances, as assessed by Levene's test for equality of variances ($p=.089$). The test indicated no statistically significant difference between the retina display group (45.69 ± 5.88) and the non-retina-display group (50.54 ± 10.35), $t(60)=.897$, $p=.374$.

Conclusion:

There was no statistically significant difference between means. The null hypothesis cannot be rejected.

A Pearson's product-moment correlation was performed to assess the relationship between performance and screen resolution in pixels-per-inch.

Result:

There was no statistically significant correlation found.

4.10.7. Comparing Effort Between Learners Grouped by Screen Resolution (H47)

Null Hypothesis:

H47: There will be no statistically significant difference in reported effort between the learners grouped by spatial resolution of device display.

Result:

An independent-samples t-test was run to determine if there were differences in effort between learners using retina display devices and non-retina-display devices. There were no extreme outliers in the data, as assessed by inspection of a boxplot. Effort scores for each group were normally distributed, as assessed by Shapiro-Wilk's test ($p > .05$), and there was homogeneity of variances, as assessed by Levene's test for equality of variances ($p = .295$). The test indicated no statistically significant difference between the retina display group (61.94 ± 6.18) and the non-retina-display group (68.12 ± 9.31), $t(60) = 1.181$, $p = .242$.

Conclusion:

There was no statistically significant difference between means. The null hypothesis cannot be rejected.

A Pearson's product-moment correlation was performed to assess the relationship between effort and screen resolution in pixels-per-inch.

Result:

There was no statistically significant correlation found.

4.10.8. Comparing Frustration Between Learners Grouped by Screen Resolution (H48)

Null Hypothesis:

H48: There will be no statistically significant difference in reported frustration between the learners grouped by spatial resolution of device display.

Result:

An independent-samples t-test was run to determine if there were differences in frustration between learners using retina display devices and non-retina-display devices. There were no extreme outliers in the data, as assessed by inspection of a boxplot. Frustration scores for each group were normally distributed, as assessed by Shapiro-Wilk's test ($p > .05$), and there was homogeneity of variances, as assessed by Levene's test for equality of variances ($p = .381$). The test indicated no statistically significant difference between the retina display group (41.83 ± 7.78) and the non-retina-display group (49.83 ± 11.37), $t(60) = 1.232$, $p = .223$.

Conclusion:

There was no statistically significant difference between means. The null hypothesis cannot be rejected.

A Pearson's product-moment correlation was performed to assess the relationship between frustration and screen resolution in pixels-per-inch.

Result:

There was no statistically significant correlation found.

4.10.9. Summary of Results for Comparisons between Learners Grouped by Screen Resolution

Text highlighted in blue refers to data from non-parametric tests and indicates the median score rather than the mean. Asterisks indicate statistically significant results ($p < .05$).

Dependent Variables	Retina Display score <i>M (SD)/Mdn (IQR)</i>	Non-Retina Display score <i>M (SD)/Mdn (IQR)</i>	p-value t-test Mann-Whitney	CI (%)	Sig.	Effect size
(H42) Pre-Post Test score (%)	25.22 (11.95)	25.92 (15.01)	0.843	<80	No	
(H41) Net Task load	58.63 (13.78)	62.50 (15.35)	0.312	<80	No	
(H43) Mental Demand	68.50 (24)	71.00 (20)	0.506	<80	No	
(H44) Physical Demand	10.50 (15)	17.50 (27)*	0.038	96	Yes	$r=.268$
(H45) Temporal Demand	62.55 (20.79)	62.04 (15.22)	0.922	<80	No	
(H46) Performance	45.69 (17.39)	50.54 (24.51)	0.374	<80	No	
(H47) Effort	61.94 (18.28)	68.12 (22.05)	0.242	<80	No	
(H48) Frustration	41.83 (23.00)	49.83 (26.94)	0.223	<80	No	

Table 4-11: Summary of results between learners grouped by spatial resolution of device display

The table shows that the learners using devices with a display having a spatial resolution with retina-discernible pixels (lower spatial resolution) reported a higher physical demand than users of devices having higher resolution displays with non-retina discernible pixels (retina-display).

4.11. SUMMARY OF ALL STATISTICALLY SIGNIFICANT RESULTS

Dependent Variables	Experimental group (n=65) score <i>M (SD)/Mdn (IQR)</i>	Control group (n=65) score <i>M (SD)/Mdn (IQR)</i>	Mann-Whitney p-value	Mann-Whitney CI (%)	t-test p-value	t-test CI (%)	Effect size
(H3) Mental Demand	70.00 (23)	76.00 (26)	0.015	98	0.045	95	$r=.214$
(H6) Performance	48.00 (33)	50.00 (44)	0.049	95	0.029	97	$r=.172$

Table 4-12: Statistically significant results between mobile-learners and non-interactive learners

Dependent Variables	Smartphone Learners (n=36) score <i>M (SD)/Mdn (IQR)</i>	Non-interactive Learners (n=65) score <i>M (SD)/Mdn (IQR)</i>	Mann-Whitney p-value	Mann-Whitney CI (%)	t-test p-value	t-test CI (%)	Effect Size
(H9) Net Task load	57.14 (13.44)	63.32 (14.80)	0.013	98	0.040	95	$d=.460$
(H11) Mental Demand	66.50 (25)	76.00 (26)	0.002	99	0.006	99	$r=.302$
(H15) Effort	62.00 (28)	72.00 (26)	0.028	97	0.014	98	$r=.219$

Table 4-13: Statistically significant results between smartphone-learners and non-interactive learners

Dependent Variables	Tablet Learners (n=24) score <i>mean rank</i>	Non-interactive Learners (n=65) score <i>mean rank</i>	Mann-Whitney p-value	Mann-Whitney CI (%)	t-test p-value	t-test CI (%)	Effect Size
(H20) Physical Demand	57.19	43.18	0.020	97	0.060	93	$r=.224$

Table 4-14: Statistically significant results between tablet-learners and non-interactive learners

Dependent Variables	Smartphone Learners (n=36) Score M (SD)/Mdn (IQR)	Tablet Learners (n=24) Score M (SD)/Mdn (IQR)	Mann-Whitney p-value	Mann-Whitney CI(%)	t-test p-value	t-test CI (%)	Effect Size
(H25) Net Task load	57.14 (13.45)	64.75 (14.92)	0.174	92	0.044	95	d=.566
(H27) Mental Demand	66.50 (24)	73.50 (14)	0.027	97	0.010	98	r=.286
(H28) Physical Demand	10.50 (15)	17.50 (25)	0.030	97	0.019	98	r=.281
(H31) Effort	60.02 (19.40)	71.00 (19.24)	0.017	98	0.036	96	d=.565

Table 4-15: Statistically significant results relating to tests between mobile-learners grouped by device type

Dependent Variables	Small (n=10) score M (SD) Mdn (IQR)	Med (n=17) score M (SD) Mdn (IQR)	Large (n=13) score M (SD) Mdn (IQR)	V large (n=20) score M (SD) Mdn (IQR)	p-value ANOVA Kruskal-Wallis	CI (%)	Effect size
(H33) Net Task load	61.78 (7.73)	57.81 (15.57)	52.96 (11.62)*	66.10 (15.81)*	0.049	95	d=1.131
(H39) Effort	65.00 (22)	50.00 (18)*	61.00 (24)	73.00 (25)*	0.034	95	$\eta^2=.143$
(H40) Frustration	37.50 (24.56)	48.65 (19.65)	29.77 (21.83)*	55.65 (25.90)*	0.014	98	d=1.186

Table 4-16: Statistically significant results between m-learners grouped by size of device display

Dependent Variables	Retina Display Mdn (IQR)	Non-Retina Display Mdn (IQR)	p-value Mann-Whitney	CI (%)	Effect size
(H44) Physical Demand	10.50 (15)	17.50 (27)	0.038	96	r=.268

Table 4-17: Statistically significant results between m-learners grouped by resolution of device display

The small differences in results between smartphone-learners in each category is due to the difference in sample size used. Some participants did not provide the make and

model of their smartphone. These data could be included in broad comparisons between smartphones and tablets, but had to be excluded from comparisons that required information about features such as screen size and spatial resolution.

4.12. SUMMARY OF CHAPTER 4

In summary, there were 14 statistically significant results from the study ($p < .05$).

Mobile-learners found the learning activity to be less mentally demanding than non-interactive learners and reported better performance in the task.

Smartphone learners reported a reduced overall task load compared to non-interactive learners with lower levels of mental demand and effort being the contributing subscales. Smartphone learners also experienced a lower net task load than tablet-learners with lower levels of mental and physical demand, and a lower effort.

Tablet-learners reported a higher level of physical demand compared to non-interactive learners. For this learning activity, tablets offered no advantage over non-interactive learning in any of the dimensions of the NASA TLX.

Learners using devices with very large screen sizes (9.7"-10.1") reported a higher task load and higher level of frustration compared to learners having large screen sizes (5.5"-9.6") and a higher level of effort compared to learners having medium screen sizes (4.7"-5.4").

Retina display device users reported a lower level of physical demand, compared to non-retina-display device users.

5. CHAPTER 05 – ANALYSIS AND DISCUSSION

In this chapter, the results of the study are evaluated in the context of the theoretical constructs of Cognitive Load Theory (CLT) (Sweller 1988) the Cognitive Theory of Multimedia Learning (Mayer 2009) and Human Computer Interaction (HCI) (Card, Newell and Moran, 1983). The intention is to draw conceptual conclusions based on connections between the findings of this study and relevant concepts from the topic literature.

This research used a quantitative approach to make comparisons between two groups of learners either undertaking a non-interactive learning activity or an interactive learning activity on a mobile device. The activity involved studying a labelled photograph of the base of a human skull and learning the names of the anatomical features found in this area.

Secondary data was collected regarding the type of device used by the mobile-learners and also the physical characteristics of the device display.

Self-reported task load was subjectively measured for both groups using the NASA Task Load Index (TLX) tool (Hart and Staveland, 1988) and objectively assessed by pre and post-testing (performance outcome measurement) (Brünken, Plass and Leutner, 2003).

The results obtained can be evaluated from various perspectives.

From the viewpoint of cautious generalisability to a larger population, results can be grouped by statistical significance, and whether the null hypothesis was rejected. As a two-tailed test, there was no assumption that results would have a particular direction. From a more informal standpoint, descriptive statistics can be used to analyse possible patterns or trends and also to suggest where further research might be useful to gain statistically significant data to support any non-inferential findings.

In this chapter, all statistically significant results will be interpreted in the light of the current literature and any non-inferential data that seems to show a pattern or trend will be examined.

5.1. EVALUATION OF PARTICIPANT DEMOGRAPHICS

To make any possible generalisations from the results of this study, it was necessary to ensure that the participants were representative of the target population. The learners for whom the study was intended to generalise are distance-learners in the field of medical education, particularly medical imaging. At the time of writing the UK health sector employed more women than men (ratio 77%:23%) (NHS Employers Association, 2015), and Diagnostic Radiographers (medical imaging personnel) having a female/male ratio of 84%:16% (Society of Radiographers, 2009). The gender ratio in distance learners is also weighted toward female students 53%:47%. Distance learners tend to be post graduates (37%) and have a mean age of 34 years (Dabbagh, 2007).

The participants recruited for this study had a mean age of 36 years, and there was a female to male ratio of 69%:31%. This proportion is approximately representative of the population of interest and was to be expected as the participants were recruited from this population by consecutive sampling. The advantages and limitations of consecutive sampling are covered on pages 127 and 156.

There were no statistically significant differences demonstrated between participants when grouped by age or gender. There was also no statistically significant difference in age or gender between groups.

5.2. EVALUATION OF FINDINGS BETWEEN GROUPS

5.2.1. Evaluation of Results Comparing Net Task load Between Groups (H1)

This test related to the primary hypothesis (H1) that there would be no statistically significant difference in net task load reported between the mobile learners and the control group during the learning activity. This hypothesis is of principal relevance because if the use of a mobile device is found to impair learning when compared to a non-interactive learning source such as a paper document (for example a book or classroom hand-out), the value of a mobile device as a learning tool may not be justified in favour of the less-expensive alternative.

In this study, it was important to control for any variables that might have contributed to extraneous cognitive load in either group but did not relate to the use of a mobile device in the learning task. The same learning activity was used for both groups to control for differences in intrinsic cognitive load due to the subject complexity (Hollender, et al., 2010). The photograph used was identical, and the labelling was the same size and used the same typeface. The activity was timed to ensure that both groups had the same opportunity to memorise the structures on the photograph.

The design of the teaching materials incorporated the principles of good instructional design as informed by HCI Theory (Hollender, et al., 2010; Zhang and Galletta, 2015). These measures included ensuring that the activity was relevant to learning and that the elements of the graphic were spatially contiguous (labels in proximity to structures on the diagram). For the mobile-learners group, redundancy effects (whereby extraneous cognitive load is increased when learners attempt to assimilate audio and visual information simultaneously) were controlled for by presenting verbal and textual cues separately. Hypothetically, any difference in extraneous cognitive load between groups would only be generated by interacting with the mobile device.

It is well documented in the literature that HCI requires a degree of visual, mental and physical coordination, all of which may increase demands on the user (Hjortskov et al., 2004). Modern mobile devices are powerful and highly complex computers and present a similar set of challenges to the user as may be found in any other type of human-

computer interaction (Kukulska-Hulme and Traxler, 2005; Deegan and Rothwell, 2010). If these demands occupy resources in short-term memory, there is likely to be a corresponding shortfall in the germane resources available for successful learning to occur (Sweller, 1994; van Merriënboer and Ayres, 2004). In relevance to this study, CLT has also been applied in the field of healthcare education and specifically in learning anatomy with labelled diagrams (van Merriënboer and Sweller, 2010).

It could, therefore, be assumed that the experimental-group would report a higher net task load than the control-group; however, there was no statistically significant difference in reported net task load between groups. On inspection of the data, it was found that contrary to the assumption made above, the result was in the other direction. The non-interactive learners reported a higher mean task load score than the mobile-learners ($M=63.32$ vs. $M=59.81$). This result was unexpected given the fact that the non-interactive learners were only required to study a labelled photograph for ten minutes, with no distraction from other sources and there was no additional activity required such as interaction during the learning activity. Conversely, the mobile-learners were required to interact with the devices and the software and might have been expected to experience some distractions and increased cognitive-load associated with device use. Although the greater mean task load score reported by the non-interactive learners is non-statistically significant, the confidence level for this result is 82%. It can be inferred from these results that mobile device interaction did not increase extraneous cognitive load in the learners and if there is any direction to the result it is possible that the use of mobile devices reduced extraneous cognitive load.

5.2.2. Evaluation of Results Comparing Net Task load Between Smartphone Learners and the Control Group (H9-H16)

As stated in the literature review in chapter 2, the intention of this study was to compare mobile devices having a ubiquity and form factor that replicates that found in textbooks. For this reason, it was decided to run the same data analysis but excluding tablet devices (H9 – H16). Tablet devices, particularly larger models, are difficult to hold or operate with one hand (Raptis, et al., 2013) and are not pocketable. They do not offer the same degree of ubiquity, portability and usability as smartphones or books. When tablets were

excluded from the mobile-learners data, there was a statistically significant difference demonstrated between the groups (with a small effect size). The non-interactive learners reported a greater level of task load than the smartphone-learners ($M=63.32$ vs. $M=57.14$).

The significant contributors to the higher levels of task load in the non-interactive group were the sub-scales of mental demand (H11) and effort (H15). A possible effect that may provide an explanation for this increase in task load is described in CLT (Sweller, 1989) and the Cognitive Theory of Multimedia Learning (Mayer, 2009). There are various conceptual models that relate to cognitive load during Human-Computer Interaction. Hollender et al. (2010) define a model where extraneous cognitive load is divided into two components, that which is induced by instructional design, and that which is induced by using the device (and software). They recognise that the aim of good educational software design is to increase germane cognitive resources (referred to as germane cognitive load by the authors) by reducing extraneous cognitive load. The first effect that might have reduced extraneous cognitive load in the mobile-learners group is the modality effect.

[The Modality effect](#)

During the learning task it was necessary to present the mobile-learners with an authentic learning activity containing the features that might be typically accessed on a mobile device. In addition to the requirement to interact with the touch-screen, the mobile app also included the addition of verbal narration, whereby the names of the structures tapped by the learners were verbally described over their headphones. Audio was not included in the control group activity as the learning task only required the participants to study a photograph in silence. This multimedia mode of delivery might have decreased cognitive load in mobile-learners group due to the modality effect introduced on page 66. In the words of Mayer (2009) this effect relies on the idea that audio and visual components of a presentation are processed by the brain separately. This assumption, which is based on the working-memory model of Baddeley and Hitch (2000) (see page 60) has been convincingly supported by the use of Functional Magnetic Resonance Imaging (fMRI) whereby cortical activation in the brain can be visualised and

shows that different areas of the brain are involved with audio and visual processing. (Buchweitz, et al., 2001.; Crottaz-Herbette, Anagnoson and Menon, 2004). The theory postulates that active learning should result in the construction of a coherent mental representation (in the case of this study, a mental representation of the base of the human skull and the names of the structures found in this anatomical area). Mayer (2005, p.38) identifies five processes that occur during this type of learning all of which are represented in Figure 5-1.

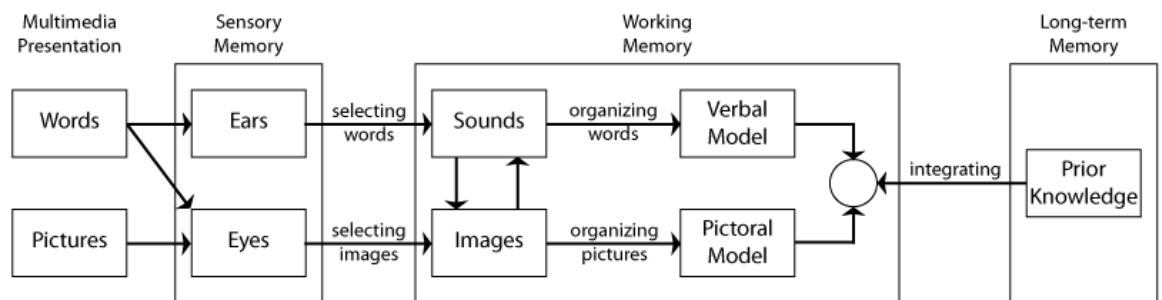


Figure 5-1: Diagram of The Cognitive Theory of Multimedia Learning (Mayer, 2005, p.37)

The learner must:

1. “select relevant words for processing in verbal working memory,
2. select relevant images for processing in visual working memory,
3. organise selected words into a verbal model,
4. organise selected images into a pictorial model,
5. integrate the verbal and pictorial representations with each other and with prior knowledge”. (Mayer, 2005, p.38)

Some events will occur simultaneously and undergo different streams of processing by the brain. This process is analogous to a broadband fibre-optic cable that permits numerous streams of data to be transported simultaneously. In this case, words are processed by the ears and eyes via the optic and auditory nerves (with printed words being mentally articulated as sounds) and the relevant sections of the photographic image being processed via the eyes and optic nerve. Visual and auditory information is

then held briefly in working memory where the learner organises the words and images to create both verbal and pictorial models of the information. These separate models are then finally integrated with prior knowledge to create a composite mental model before being consigned to long term memory as a schema. Mayer (2005) goes on to suggest that multimedia presentations that are designed to take advantage of the way the brain processes audio and visual information are therefore less likely to overload the buffer of the short-term memory, and in doing so reduce what Sweller identifies as extraneous cognitive load.

There is, however, some more recent research that goes against this proposition. Savoji, Hassanabadi and Fasihipour (2011) conducted a between-groups study looking at multimedia presentations (on laptop computers) with varying degrees of narration. Their results (a self-reported mental-effort scale) did not support the hypothesis that verbal information presented as narration instead of on-screen text will improve learning outcomes. However, the authors indicated that the learning task in question was user-paced, and this factor is known to reduce the modality effect because students have time to assimilate both written text and imagery. This effect is well documented in the literature and was identified in a meta-analysis of the modality effect conducted by Ginns (2005). Ginns established that the modality effect is moderated by the pacing of the presentation and students undertaking system-paced activities (across a wide range of topics) benefitted more than students who had time to assimilate the textual element of the learning materials at their own user-defined pace.

My study used a timed learning task; both groups had ten minutes in which to learn from the activity and a ten-minute timer was displayed to both groups. The activity was not user-paced, but system-paced. The non-interactive learners would have been required to split their attention between the graphic element of the photograph and the labels in forming an integrated mental model (see Figure 5-2). Furthermore, the mobile app was designed to present the textual information in addition to the audio narration but timed to appear a few seconds after the audio description. This would have provided additional input into what Mayer (2005) describes as “sensory memory”. This design feature is also likely to have resulted in a reduction of extraneous cognitive load because the timing of the audio component was designed to ameliorate redundancy effects (defined on page 67).

Split-Attention Effect

The mobile app was also configured to position the text closer to the structure being described than was possible on the non-interactive photograph used for the control group.

Close spacing of text labels is another feature of good HCI and Instructional Design Theory as it provides what is known as high spatial contiguity (Mayer, 2009). Closely spaced elements of information are less likely to cause split-attention, identified by Ayres and Sweller, 2005) as a contributor to extraneous cognitive load.

Moreover, the m-learners were not required to split their attention between the text and graphic elements of the learning materials because the verbal narration would have allowed them to identify the structures on the diagram without the necessity to divert attention to the text labels.

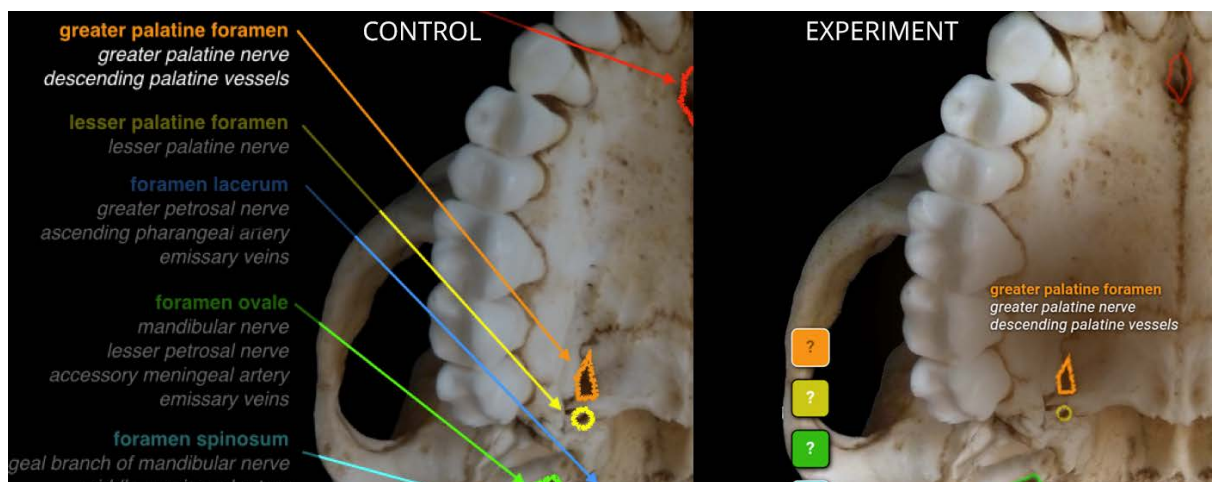


Figure 5-2: Spatial contiguity - The mobile app (right) permitted the labels to be spaced close to the structures on the photograph

It is proposed, that in the case of the mobile-learners group, any extraneous cognitive load that might have been imposed by having to use the device has been ameliorated by the modality effect and the contiguity principle both of which were considered in the multimedia design of the learning materials.

User Engagement

A third plausible explanation relates to the interaction required with the touch-screen of the device and how this might foster user-engagement. Over the last five years, student engagement has become a focus of research in the field of education. The Higher Education Academy recently reported that the evidence shows a convincing correlation between high levels of engagement and successful learning outcomes (Trowler and Trowler, 2010). The nature of student engagement is described by Kahu (2013) as being multi-faceted with the psychological perspective being cited as one of the main fields of research (in addition to behavioural and psycho-social processes), cognition being an important prerequisite for engagement. In the chapter 2 literature review, there were examples provided where mobile devices encouraged engagement through psycho-social processes. These included the use of social networking apps such as Twitter and communication afforded by Skype (Microsoft Inc.) and Short Message Service texts (Manuguerra and Petocz, 2011; Martin and Ertzberger, 2013). In the context of my study, these psycho-social dimensions do not apply, because the learners were undertaking the task individually with no collaboration. However, the answer might lie in the nature of the devices themselves and the structure provided to the learning activity by the interaction with the learning materials via the touch-screen. An interesting rationalisation for user-engagement with mobile devices is made by Sung and Mayer (2013, p.642) who suggest that “Media Equation Theory” could be responsible. This theory, conceptualised by Reeves and Nass (1996) looked at the relationship between humans and computers. They rationalised that humans were inherently polite to computers because of the way a computer was able to communicate and interact with a human being. Even in 1996, this communication was close enough to a human interaction to elicit a social response from the user. Sung and Mayer expand on the theory by suggesting that any conditions that minimise reminders that computers are not human will increase this effect. Mobile devices might, therefore, increase the sense of “social partnership” in comparison to a desktop computer - that cannot be moved or held - and in doing so “motivate learners to want to continue the relationship”.

From an educational-theory perspective, there may be an additional explanation.

Returning to Gagné’s (1985) Conditions of Learning - introduced on page 16 - there is a focus on Intellectual Skills and Cognitive Strategy. In Gagné’s model, he emphasises that

the first stage in any learning activity, particularly in terms of instructional design, is to gain the attention of learners. Mobile devices such as the iPad have been found to grab the attention of learners and foster engagement throughout a learning activity. (Liu, Li and Carlsson, 2010; Manuguerra and Petocz, 2011, Rossing, et al., 2012; Gikas and Grant, 2013, Hwang, Yang and Wang, 2013; Sung and Mayer, 2013; Chan, et al., 2014)

In the study by Rossing et al. (2012) students commented that iPads caused them to pay more attention during class, that the learning activities kept their attention and kept them involved. Manuguerra and Petocz (2011) also reported a higher level of student engagement – even with traditionally “dry” subjects and that attrition rates dropped dramatically in the cohort that used iPads to augment their learning.

In my study, the control group were asked to study a photograph. There was no interaction required during the learning activity and nothing other than the labelled photograph to hold their attention over the ten-minute task. It is therefore proposed that the mobile devices, due to their novelty and the necessity for the students to interact, offered a more engaging learning experience to the experimental group, and this was reflected in their lower reported task load. In the words of Mayer (2014, p. 173) the “motivational features” of the multimedia presentation “improved student learning by fostering generative processing”. Generative processing being defined as the cognitive processes relating to the learners effort and engagement during the learning process in the task of meaning-making.

5.2.3. Evaluation of Results Comparing Task load between Tablet Learners and the Control Group (H17-H24)

For completeness a t-test was also used to compare the task load between tablet-learners and non-interactive learners. There was no statistically significant difference in task load demonstrated (*tablet-learners* $M=65.15$ vs. *non-interactive* $M=63.32$).

Furthermore there was no difference between tablet-learners and the control group in any of the subscales of the NASA TLX, with the exception of physical demand. Tablet learners found the task to be more physically demanding than the control group ($Mdn=14$ vs. $Mdn=18$). In short, a tablet computer offered no advantages over a textbook in this learning activity.

5.2.4. Evaluation of Results Comparing Test Scores between Groups (H2)

The null hypothesis here (H2) was that there would be no difference in pre/post-test scores between the mobile learners and the control group. The pre/post-test was included in the research design as an indirect, objective measurement tool for cognitive load (as a performance outcome measure) and also to measure any difference in learning outcomes. As a performance outcome measure, it can also be used to validate the results of the NASA TLX test. The assumption being that any statistically significant difference between groups, or direction of results from the self-reported task load scores will be reflected in the pre/post test results. The results of the pre/post test score supported the results from the NASA TLX in showing no statistically significant difference between groups. On examination of the data, it can be seen that the median score for the mobile-learners was 26.50 compared to the non-interactive learners whose median score was 17.50.

The mean scores more accurately reflect the average marks that would have been awarded to both groups in a summative assessment and these were 25.3 for the mobile-learners and 23.6 for the non-interactive learners. These results were not statistically significant ($p > .05$), it can only be stated with 85% confidence that the difference is due to the independent variable. However, the direction of the result may provide further support for the NASA TLX result as a lower pre/post-test score would indirectly confirm an increase in cognitive load as shown by a greater net test-score. Although the differences in NASA TLX *net* test score were not statistically significant between groups (when tablet computers were included), this apparent correlation between pre/post test score and performance can be explored in more depth by examining the individual sub-scales of the NASA TLX. These are covered below.

5.2.5. Evaluation of Results Comparing Mental Demand Between Groups (H3)

Hypothesis H3 stated that there would be no statistically significant difference in mental demand reported between the mobile learners and the control group.

As a multi-dimensional scale, the NASA TLX asks participants to report on their perceived level of mental demand relating to the learning task. When this dimension was examined in isolation, there was a statistically significant difference between groups ($p=.015$). The non-interactive learners found the learning activity more mentally demanding than the mobile-learners group ($M=76$ vs. $M=70$). The effect size was small (increasing to medium when the tablet data was excluded), but this result offers an explanation for the direction of the net task load result. As discussed on page 22, Mental demand is affected by the number of elements that must be simultaneously processed by the working memory (Miller, 1956; Cowan, 2001; Arsalidou, et al., 2013). Task complexity also plays a role, van Merriënboer and Sweller (2010) state that intrinsic cognitive load is affected by task complexity and the expertise of the learner. In my study, the expertise-reversal effect (described on page 63) was controlled by choosing a “difficult” learning topic (evidenced by the poor pre-test scores) in which none of the participants were likely to have pre-existing schemata. This leaves the complexity of the learning task as the primary contributor to intrinsic cognitive load and offers a possible explanation for the difference between groups. Figure 5-2 shows that the mobile-learners were only concentrating on one element of the diagram at any given moment. The mobile app software was designed to display only the information that was relevant to the structure tapped by the learner. Labels and information relating to the other structures on the diagram were hidden from view until activated. Conversely, the non-interactive learners were presented with a photograph that featured relatively dense, complex labelling that was necessary to provide the same information as the interactive version of the learning materials. van Merriënboer (2010) states that when learning anatomy, intrinsic cognitive load can be managed by using learning tasks that firstly present isolated elements (known as having “low element interactivity”) and then work up to full complexity. This is how the mobile app was designed. By reducing intrinsic cognitive load, the mobile-learners would have

more germane resources available for learning and would therefore find the task less mentally-demanding than the non-interactive group.

5.2.6. Evaluation of Results Comparing Performance Between Groups (H6)

It was hypothesised that there would be no statistically significant difference in task-related performance reported between the experimental group and the control group (H6). Participants were asked to rate their own (perceived) performance using a Likert scale where 0 represents perfect performance, and 100 represents a complete failure of the task. Scores for the non-interactive group reported a significantly poorer performance (mean rank=50) than for the mobile-learners (Mean Rank=48). Poor performance is of relevance because it may indicate that the task load has become too high (Chen, et. al, 2016).

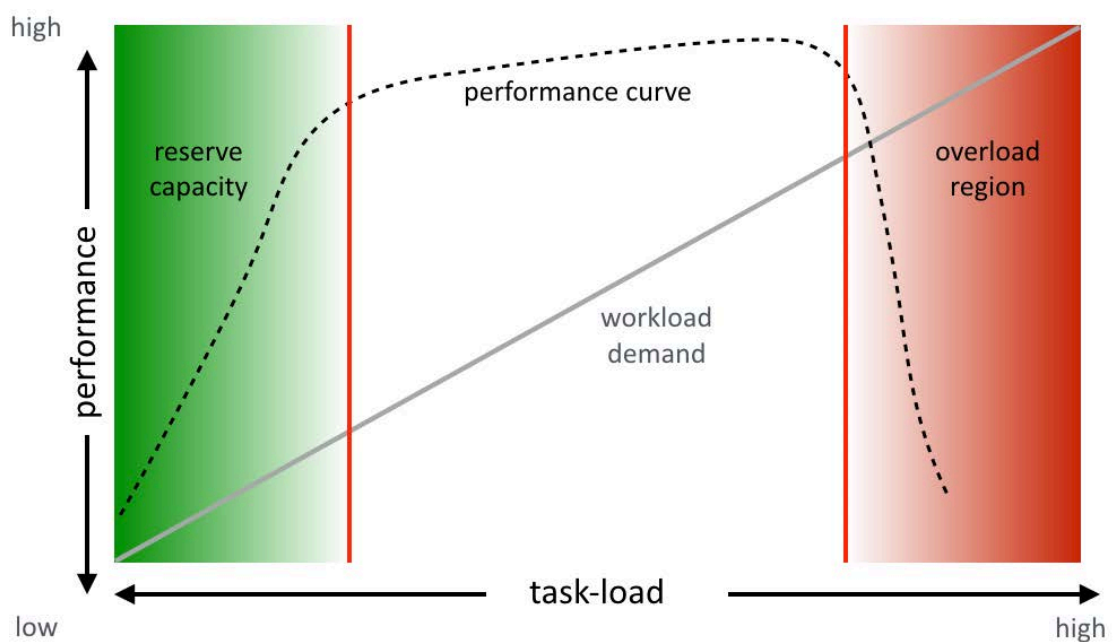


Figure 5-3: Performance vs. task load - adapted from Chen, et. al. (2016, p.39)

However, as can be seen in Figure 5-3, performance *increases* with task load, until a certain threshold is reached. The critical point is where the germane resources of working memory (what the diagram refers to as “reserve capacity”) are exceeded by the demands

of the task (van Merriënboer and Sweller, 2010). According to Chen, et al. (2016), a difference in performance could, therefore, be due to one of two factors:

- a) the non-interactive learners experienced a lower net task load than the mobile learners
- b) the non-interactive learners reached the cognitive overload region

The factor that is more likely to be the explanation in this case, can be determined by examining the reported mental demand for both groups. If the cause of the poorer perceived performance were due to a lower task load, it would be expected that the reported mental demand would also be lower for the interactive group, as they would still have more germane resources in working memory compared to the mobile-learners. This was not the case. The non-interactive learners reported a statistically significantly higher mental demand than the mobile-learners. The difference in reported performance between groups is, therefore, likely to be due to the non-interactive learning activity causing a greater cognitive overload than the m-learning activity. This is the section of the performance curve where the demands of the task exceed the germane resources in working memory shown in red in Figure 5-3. This result should be regarded with caution however as the effect size was small.

On page 249 it was stated that the pre/post test score result could be used as a validation tool for the net task load result as a performance outcome measure. It can also be used to validate the performance sub-scale of the NASA TLX.

In addition to the statistically significant difference between groups, there is also a statistically significant (weak) positive correlation between increasing overall performance and pre/post test score. This supports the use of the pre/post test score as a performance outcome measure.

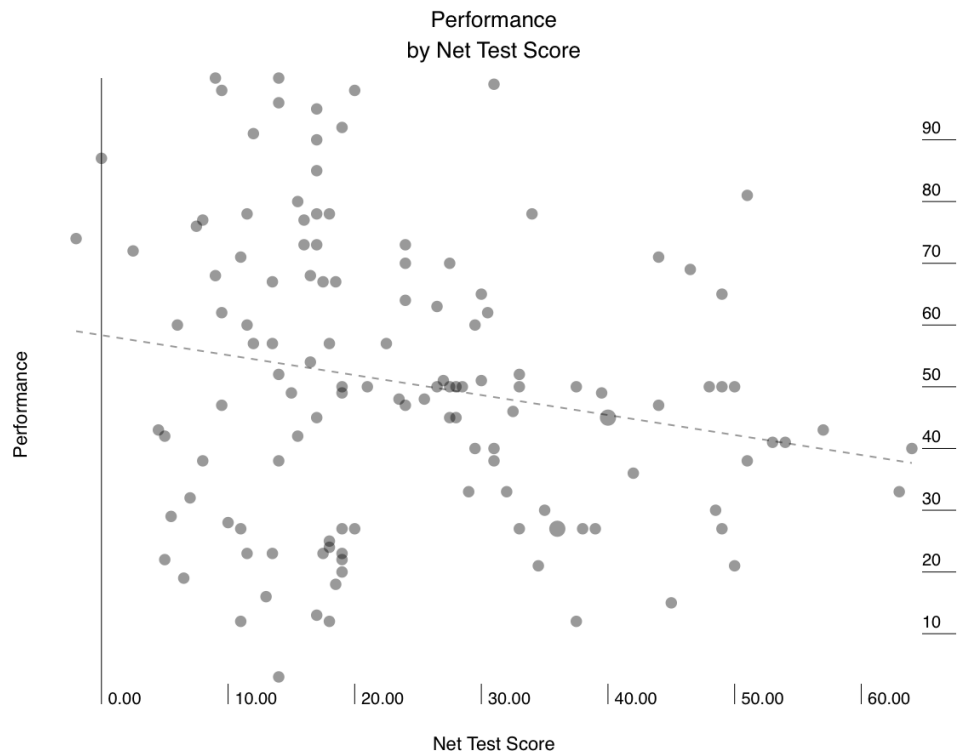


Figure 5-4: Scatterplot showing a weak correlation between performance and pre/post-test-score

The correlation in Figure 5-4 appears to be negative because of the way performance is scored in the NASA TLX, a perfect performance has a score of zero. This result does not necessarily indicate a causal relationship because other elements of the activity could have affected both performance and pre/post-test score. But it can be stated that the group who reported a poorer average performance were correct to do so.

5.2.7. Evaluation of Results Comparing Physical Demand Between Groups (H4)

The mobile-learners reported a (non-significant) higher physical demand score than the non-interactive learners as might be expected. The non-interactive group were not required to perform any physical activity at all during the learning task other than to study a photograph. Although this difference was not statistically significant, there were statistically significant differences in physical demand between learners when grouped by device. These are evaluated in a following section on page 259.

5.2.8. Evaluation of Results Comparing Temporal Demand Between Groups (H5)

The data suggests that mobile-learners felt that they were under more time pressure than the non-interactive learners evidenced by a higher level of temporal demand ($M=61.86$ vs $M=56.03$). Although the result was not statistically significant, there is an 88% confidence level that the result is due to the independent variable rather than chance. Both groups had exactly the same amount of time to complete the learning activity which suggests that if there was a genuine increase in temporal demand in the experimental group, it might be down to a difference in the subjective experience of time. In simple terms, time must have appeared to pass more quickly to the mobile-learners. There are various psychological models relating to factors that may affect the subjective estimation of time, some of which relate to cognitive load. Tsao, et al. (1983) demonstrated that if a group of participants is required to undertake an activity or stimulus that increases mental load, the duration of a secondary event is underestimated. This effect has been confirmed in more recent experiments involving tasks such as mental arithmetic, visual searching and visual tracking of an object (Brown, 1997). In this case, the non-interactive learning task (that increased mental demand) should have been estimated to pass more quickly. However, this was not the case, the mobile-learners reported a lower mental demand than the non-interactive learners, but indicated that they felt under more time pressure. These conflicting findings suggest that this non-statistically significant result is either due to chance, or given the confidence level, a

confounding variable that has not been identified. If this is the case, further research might be warranted.

5.2.9. Evaluation of Results Comparing Effort Between Groups (H7)

Effort is defined by the NASA TLX as a measure of how hard the participants had to work (mentally and physically) to accomplish their level of performance in the task. The non-interactive group reported a higher mean score for effort than the mobile-learners at a confidence level of 82% ($Mdn=65$ vs. $Mdn=72$). Although not statistically significant, this score appears to support the assertion that non-interactive learners experienced a greater level of cognitive load than the mobile learners, as mental effort contributes to this dimension and is an indicator of cognitive load (Paas and van Merriënboer, 1993). A Spearman's correlation test indicated that there was a statistically significant correlation between effort and mental demand for both groups.

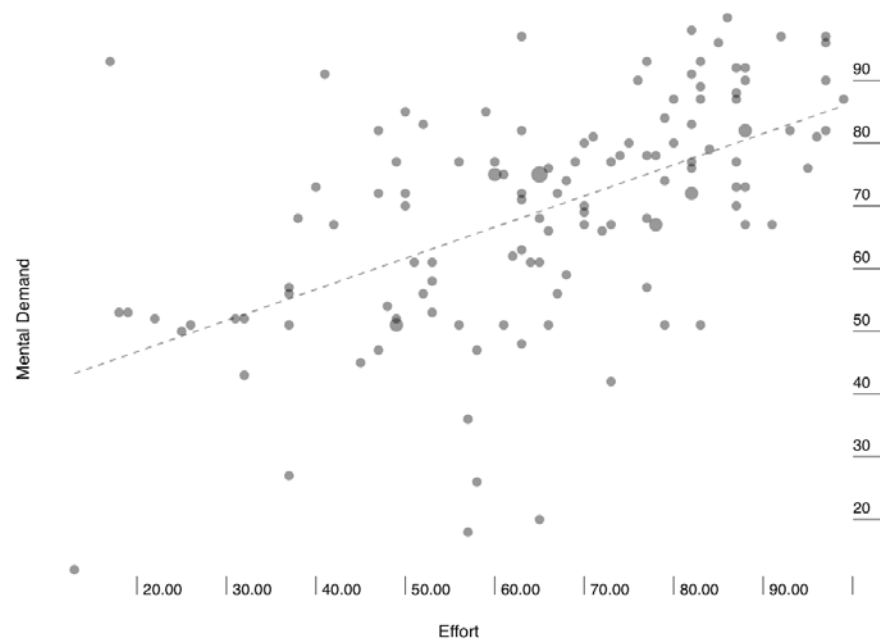


Figure 5-5: Scatterplot showing a moderate positive correlation between mental demand and effort.

There was no correlation between effort and physical demand for either group. This correlation strongly suggests that mental demand was the main contributor to the overall

effort score (rather than physical demand). Although not quite the same construct as mental demand, mental effort can be defined as “the total amount of controlled cognitive processing in which a subject is engaged. Measures of mental effort can provide information on the cognitive costs of learning, performance, or both” (Paas and van Merriënboer 1993). It is likely that the difference in effort between groups reflects the difference in mental effort (mental activity) between groups rather than differences in physical demand. This result, therefore, supports the proposition that there is an increased cognitive load experienced by the non-interactive learners.

5.2.10. Evaluation of Results Comparing Frustration Between Groups (H8)

The difference in mean rank frustration score was not statistically significant between the groups (experimental group *Mean Rank*=68.85 vs. control group *Mean Rank*=62.15). This is in contradiction to some studies from the topic literature. Demouy and Kukulska-Hulme (2010) reported frustration in mobile learners undertaking learning activities on smartphones when compared to DVD-ROM. The reasons presented in their study included limited device functionality, poor sound quality and difficulty in navigation. Other papers have cited network connectivity problems (Kukulska-Hume, et al., (2009), application (software) failure, small keyboard size, and device distraction (Gikas and Grant, 2013). Most of these potential sources of frustration were controlled in my study. The devices were typically modern and equipped with high-quality audio functionality. Scrolling navigation was minimised by careful interface design. Network connection issues were avoided by allowing pre-download of the app prior to the learning activity. Software failure was avoided by robust application build; there was no requirement to use a keyboard, and the app was designed to occupy the entire screen to reduce sources of distraction. All of these measures fall under the HCI recommendations for good user-centred software design (Hollender at al., 2010) and are likely to have reduced frustration in the m-learners group to a level that was not significantly different from the non-interactive learners.

5.2.11. Summary of the Evaluation of Principal Findings

From the results there appears to be an evidence-based outcome that there was no significant difference in task load between groups other than in the NASA TLX subscales of mental demand and reported task-performance. This strongly suggests that the mobile device and the app-based learning materials did not place any additional cognitive burden on the learners in comparison to the control activity of studying a labelled photograph.

When the statistically significant results relating to reduced mental demand and performance are considered in combination with the results of the other subscales of the NASA TLX it appears that the direction of the result lies in the other direction.

In addition to the above, when data relating to tablet computers were removed from the analysis, there was a statistically significant difference in mean task load between groups with a small-to-medium effect size. This result suggests that smartphones offer a slight advantage over tablet computers in a learning activity such as the one used in the study.

5.3. EVALUATION OF SECONDARY FINDINGS

In addition to the primary research question, the data collected allowed an investigation into differences between learners when grouped by device type. It was considered at the beginning of the study, that if the use of touch-screen devices had been shown to increase cognitive load in the users, or presented a barrier to learning, it would have been useful to know whether smartphones or tablets were the bigger contributor to these factors. In light of the previous results where smartphone-learners experienced a statistically significantly reduced task load, reported a significantly higher performance and were shown to have reduced mental demand, these results have a different implication, namely which type of device, if any, is better at reducing cognitive demand in learners across the sub-scales of the NASA TLX. These secondary findings relate to the mobile-learners group only, and as such used a smaller sample size ($n=60$). This reduces the statistical power of the data and the ability to make inferences from the results. However, the sample size is comparable to (or larger than) some previous studies in this field (Raptis, et al., 2013; Chu, 2014; Molina, et al., 2014; Lin, Wang and Kang, 2015). There is a full explanation of how potential threats to data validity were addressed in the limitations section on page 292.

There were four statistically significant results obtained that showed differences between smartphone-learners and tablet-learners. Smartphone-learners reported a lower net task load score, reduced mental demand, reduced physical demand, and lower overall effort compared to tablet-learners. These results are evaluated below.

5.4. EVALUATION OF RESULTS BETWEEN SMARTPHONE AND TABLET LEARNERS

5.4.1. Evaluation of Results Relating to Task load Between Smartphone and Tablet Learners (H25-H32)

The reported net task load was significantly greater for the tablet-learners group ($M=64.75$) than for the smartphone-learners group ($M=57.14$) with a medium effect size. This difference in task load was not supported by the results relating to pre/post test score which (as a performance outcome test for cognitive load) showed no statistically significant difference between device users. This suggests that mental demand may not have been the primary contributor to the net task load score. Mental demand (H27) was higher in the tablet-learners than the smartphone-learners ($Mdn=73.5$ vs. $Mdn=66.5$) but on further inspection of the data, it was found that there was also a higher physical demand (H28) reported by the tablet-learners ($Mdn=17.5$ vs. $Mdn=10.5$), This was confirmed by the score for overall effort (physical and mental) which was also greater in the tablet-learners. Previous studies have shown that screen size can have an effect on performance in learning tasks. Raptis et al. (2013) reported a significant increase in task-time for users with smaller screens and this finding was also confirmed in a 2014 study by Molina, et al. Furthermore, Kim and Kim (2012) found that there was an improvement in retention of information when learners used a larger screen size for a language learning-task. All of these previous findings appear to contradict the findings of this study. As the effect size of this result was medium, it was thought appropriate to determine whether there were any theoretical explanations for the apparent anomaly. There are three theory-based explanations that can be offered, these are evaluated below.

Familiarity and Preferred Screen Size

Due to the way the participants were randomised for this part of my study, some smartphone-learners were offered tablet devices to use instead of their own smartphone. Most of the participants owned smartphones and brought them to the data collection activities. However, to facilitate random allocation to smartphone and tablet groups it was sometimes necessary to provide a tablet computer to a participant who did not own

one, or failed to bring one. From an HCI perspective, user non-familiarity may present a confounding variable that had not been taken into account during the design of the data collection tool; namely that users may have a screen size preference and a familiarity with their own device. According to Raptis, et al. (2013) these factors have a strong effect on perceived device usability. Although their study looked at information retrieval rather than learning, the authors found that a significantly higher device usability rating was reported by those participants who were using their own device, or a device having their preferred screen-size and that the preferred screen-size was 4.0"-4.6" (i.e. a medium size rather than large). It was therefore considered that the reason for the higher scores reported for mental demand, and effort might have been due to the fact that the tablet-learners were interacting with devices that they were not familiar with, were not their own, or were not their own preferred screen size. Psychological comfort with a device is cited by Koole and Ally (2009) as part of their FRAME model, evaluated on page 53. Koole and Ally state that psychological comfort is a variable that both reduces cognitive load, and increases device usability regarding the speed at which tasks can be performed. The necessity to allow participants to use their own devices in research studies is echoed by Demouy and Kukulska Hulme (2010, p.218) who state that "Current thinking suggests that it is advisable, where possible, to use devices that most learners own already".

It is unlikely that device-unfamiliarity had a major bearing on my results because when the same statistical tests were run post hoc after having excluded the participants using unfamiliar devices, there was no change in the statistical significance of the results for any of the subscales of the NASA TLX or the pre/post-test results. However, device-familiarity is recognised as a potential confounding factor and is a variable that should be taken into account in future research.

Overall Device Size

As screen-size increases there is necessarily a corresponding increase in device-size to accommodate the display. There is also an increase in weight because a large display draws more power than a small display and the battery must therefore be scaled up (Carroll and Heiser, 2010; Mittal, Kansal and Chandra, 2012). A larger, heavier device becomes harder to handle. Raptis, et al. (2013) noted that task efficiency on mobile devices increased when using devices over 4.3 inches, but presented a counter argument

that device-usability decreases when the size becomes too large for the user to hold in one hand. One-handed operation is not easily accomplished with most large tablet computers and is likely to offer an explanation for the increase in physical demand in tablet-learners in this study. This is in agreement with a recent study by Pereira et al. (2015) who found that large tablet computers increased user fatigue, and reduced overall usability and productivity in comparison to smaller devices.

[Spatial Resolution of Device Display](#)

The tablet devices used in this study had a lower mean spatial resolution in pixels per inch than the smartphones. (tablets $M=175\text{PPI}$, smartphones $M=370\text{PPI}$). This is a factor that can affect cognitive load as it impacts on the performance of visuospatial short-term memory (Lin, Wang and Kang, 2015). In the absence of any other significantly different variables between tablet devices and smartphones (other than physical size and weight) it is proffered that spatial resolution may also offer an explanation for differences in reported mental demand between smartphone-learners and tablet-learners. A more detailed evaluation of device display size is presented in the following section.

5.5. EVALUATION OF RESULTS BETWEEN LEARNERS GROUPED BY SCREEN SIZE OF DEVICE USED

These tests used smaller group sizes than the previous experiments as they represent subdivisions of the mobile-learners group and only used data from participants who identified the device used ($n=60$). Inferences drawn should, therefore, be viewed with a degree of caution.

5.5.1. Evaluation of Results Relating to Net Task load vs. Screen Size (H33)

As can be seen in Figure 4-38, the smartphone results (3.5" – 5.5") show a non-statistically significant increase in net task load with decreasing screen size. However, there is a statistically significant difference in task load between large-screen smartphones and tablets with a large effect size. The tablet-learners (9.7"- 10.1") reported a higher task load than the smartphone (phablet) learners in the 5.5"- 9.6" group, notwithstanding the fact that this smartphone group were using devices with the largest smartphone screen-size. The two sub-scales that contributed to the net task load score were Effort (H39) and Frustration (H40). The tablet-learners reported significantly higher levels of frustration than the large-screen smartphone-learners during the learning task and also higher levels of effort than smartphone-learners. A post hoc test revealed that effort can be seen to correlate positively with mental demand, but not physical demand.

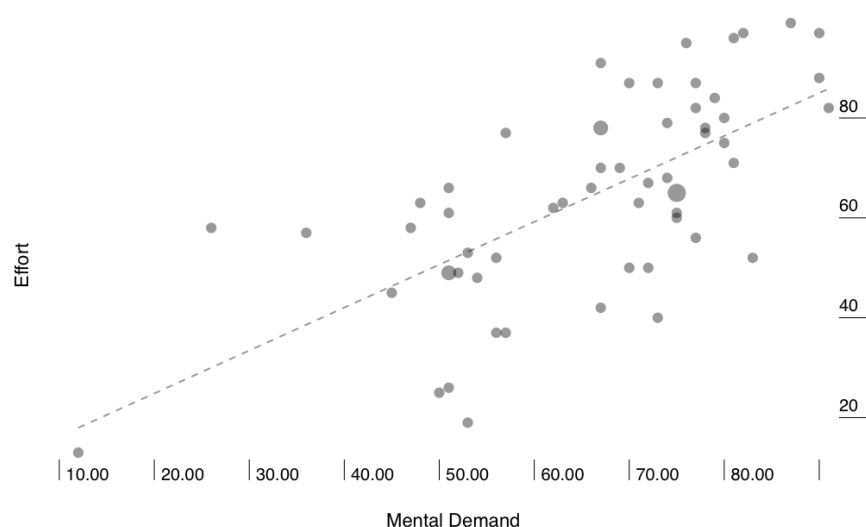


Figure 5-6: Scatterplot showing a moderate correlation between effort and mental demand in the mobile-learners group

This correlation demonstrates that mental effort was the main contributor to the overall effort score. Increasing mental effort is typically indicative of increasing cognitive load for the reasons given on page 255. These outcomes are almost entirely in agreement with the findings of Raptis et al. (2013) who found that task efficiency when undertaking information-seeking exercises increased with screen size, becoming more efficient with screen sizes over 4.3 inches. The screen-size results also support their statement concerning a limit to device size, namely that larger devices will reduce device usability because device operation will require the use of more than one hand. This provides a possible explanation for why the tablet-learners reported a higher task load than the smartphone-learners.

The relationship between cognitive load and screen size should be viewed cautiously because other variables may coincidentally correlate with screen size. There is a trend in increasing smartphone screen size that has a positive correlation with device release date. Many other features of mobile devices also have a positive correlation with release date because manufacturers understandably tend to improve the specifications of the device over time.

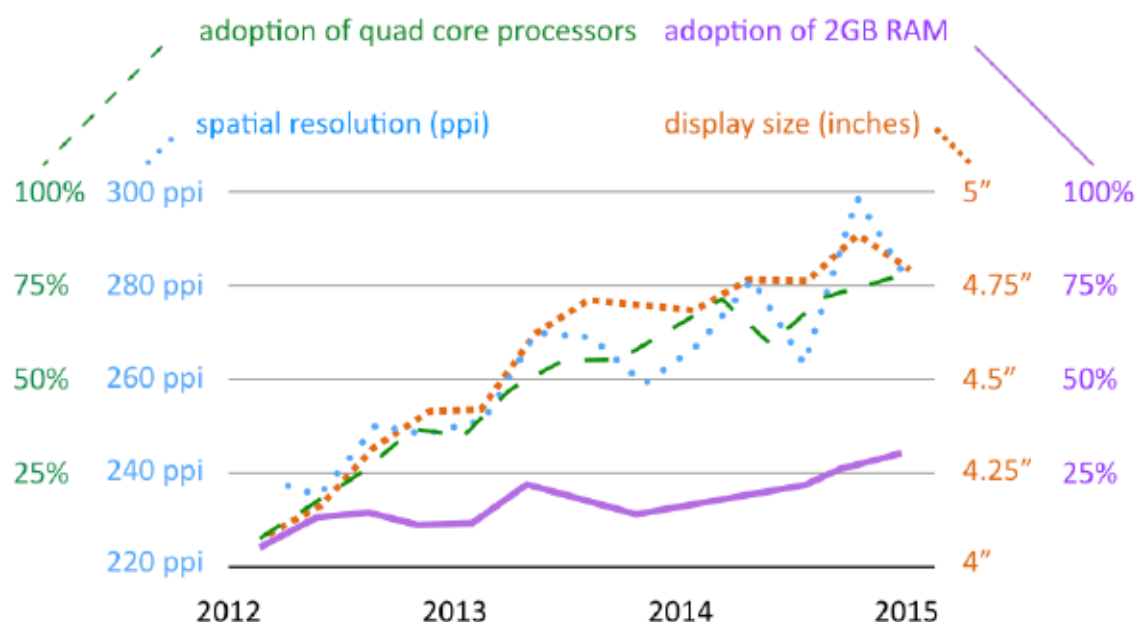


Figure 5-7: Device feature improvements by release date (Han and Cho, 2016)

As can be seen in Figure 5-7, this trend of increasing screen size over time means that, with very few exceptions, smartphone-learners having devices with larger screens are coincidentally likely to have the benefit of faster central processing units (CPU), a larger random-access memory capacity (RAM) and improved spatial resolution display. These features result in devices that are faster, are capable of running more sophisticated applications and have a clearer display where individual pixels are no longer discernible by the human retina (Han and Cho, 2016). Whether or not these factors would have had an effect on task load is difficult to determine. Benchmarking processor speed is a complex task because there are many variables and different scales are required to assess performance on different tasks such as web browsing, gaming or intense 3D graphic display. It was difficult to find comparative data across all of the 32 different device models used in this study, and there is also a risk that some participants might have wrongly identified the particular model used (for example, there are at least 18 versions of the Samsung Galaxy tablet available). For these reasons, a comparison between CPU benchmarks and task load was beyond the scope of this study. There may, however, be a possible agenda for further research into this area. It is unlikely that CPU speed will have had any significant effect on the results of this study because the learning materials were designed to be non-processor-intensive. Responsiveness of interactivity would not have differed significantly between devices. Furthermore, had the graphics processing slowed down the learning activity, it would be expected that the reported temporal load would show a corresponding difference between users of different devices, and this was not the case (see Figure 4-42).

Regarding the spatial resolution of the device display, this data was readily available from the manufacturers and results relating to display resolution are evaluated later in section 5.6.

5.5.2. Evaluation of Results Relating to Effort and Screen Size (H39)

In the subscale of effort, the learners using very large-screen devices (9.7" - 10.1") reported a significantly higher level of effort than the medium-sized screen users ($Mdn=70$ vs. $Mdn=53$) the effect size was medium. Analysis of the subscales of the NASA TLX showed a statistically significant positive correlation between effort and mental demand and an explanation for this was provided in section 5.4.1.

5.5.3. Evaluation of Results Relating to Frustration and Screen Size (H40)

The results revealed that the increase in user frustration between a large ("phablet") to very large (tablet) screen size was statistically significant ($M=29.7$ vs. $M=55.65$). This was the only statistically significant result relating to user frustration. The difference in frustration reported between smartphone-learners (of various screen sizes) was not statistically significant, but there was a statistically significant between phablet learners having large screens (5.5" - 9.6") and tablet-learners (9.7"-10.1"). This seems to represent a somewhat subtle difference in device form-factor (for example the difference between an iPad mini, and a standard iPad) the result was significant to $p=.014$ and with a large effect size. There are two possible explanations offered.

Device Size

The size increment between large smartphones and tablets represents the point where a device becomes too large for single-handed operation. Usability is therefore decreased as previously stated by Raptis et al. (2013) and Pereira et al. (2015) leading to user frustration.

Spatial Resolution

A lower (mean) spatial resolution in pixels-per-inch was also exhibited by the tablets when compared to smartphones. The largest difference was between large screen smartphones (mean PPI=313) and tablets (mean PPI=172). This lower resolution is likely to have increased the necessity for zooming and scrolling of content by the user when attempting to identify small structures on the image. It is suggested that this is likely to have caused a degree of frustration in these participants. Sub-optimal spatial resolution

may also adversely affect cognitive load. In a learning activity that relates to anatomy, a high spatial resolution will provide a sharper image. The viewer is more easily able to resolve small structures such as the ones used in the experimental learning task. This improved spatial resolution is likely to aid the learner, because, as stated on pages 60 and 243, the working-memory model of Baddeley and Hitch (1974), and the Cognitive Theory of Multimedia learning both feature visual input. Previous research into this area by Lin, Wang and Kang (2015) has shown that inferior spatial resolution in an image causes a corresponding decrease in visual acuity, visual identification and correspondingly a decrease in visuospatial memory performance – a factor that is known to impede schemata-formation.

5.6. EVALUATION OF RESULTS FOR COMPARISONS BETWEEN LEARNERS GROUPED BY SPATIAL RESOLUTION OF DEVICE DISPLAY

There appear to be no studies presented in the current topic literature that examine the effect of spatial resolution on learning outcomes for mobile-learners. Nonetheless, high screen resolution is a feature of mobile devices that is emphasised by the manufacturers as being important - the so-called “retina-display”. In the context of this study, the spatial resolution of the mobile device is an important variable because the intention was to compare mobile devices to textbooks. Lau and Arce (2008) explain that textbooks typically feature halftone-printing that uses print-screens having a certain number of lines (of dots) per inch (lpi). The aim with printed graphics is to use a resolution that renders the dots indiscernible to the naked eye; this threshold is approximately 200 lpi. In Figure 5-8 it can be seen that some of the smaller details cannot be discerned as easily on the 100 lpi image compared to the higher spatial resolution 200 lpi image.

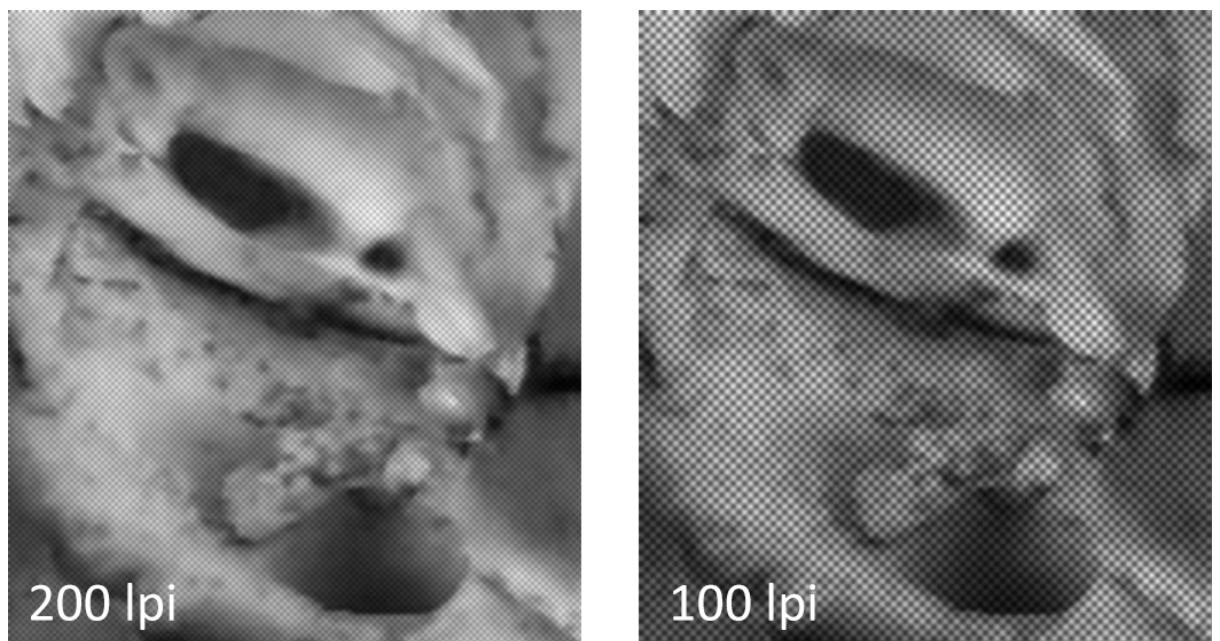


Figure 5-8: Resolution comparison in halftone printing (magnified)

Most of the major smartphone manufacturers now offer models with a high pixel density, usually presented as pixels per inch. As in printing, the aim of using a high pixel-density display is to produce an image with high spatial resolution. The reason that this is an

important feature in mobile devices relates to the fact that they typically have a shorter viewing distance than a computer screen. The reason for providing a high resolution display is that a person with average visual acuity will not be able to see individual pixels at a viewing distance of 30cm. Spencer, et al. (2013) state that the widely accepted visual acuity cut-off value for 20/20 vision is equivalent to viewing a display having a pixel density of 290 PPI, but that individuals with sharper visual acuity may be able to distinguish images having a higher resolution.



Figure 5-9: Comparative spatial screen-resolution (magnified) of 400 PPI (left) and 100 PPI (right)

Results relating to spatial resolution are evaluated here from two perspectives.

The first data analysis looks at differences between learners grouped by device resolution, namely a “retina-display” group who used devices having a spatial resolution of greater than 290 PPI (not distinguishable by the human eye) and a “non-retina-display” group who used devices having a spatial resolution of fewer than 290 PPI (distinguishable by the human eye – see Figure 5-9 above) .

The second data analysis looks at correlations between screen resolution in pixels-per-inch and the sub-scales of the NASA TLX.

5.6.1. Evaluation of Results Relating to Differences between Retina-Display vs. Non-Retina-Display Devices

The aim of this test was to determine whether the learners having a display with retina-discernible pixels reported a difference in cognitive load when compared to learners whose devices has non-discernible pixels.

The only variable that appeared to be affected by the spatial resolution of the device display was the subscale of *physical demand*. This factor was statistically significantly greater in learners using non-retina-display devices ($Mdn=17.50$) than in learners using retina-display devices ($Mdn=10.50$). This result appears to be counter-intuitive. Spatial resolution has a bearing on how well the device can display detail on the image. However, it is difficult to conceive how spatial resolution would affect *physical* demand. The effect size of this result is small, but as it has statistical significance an evaluation is offered. There are two possible explanations for this result, physical demand due to increased interaction with the device (pinching/zooming) or due to the physical device size.

Increased Necessity for Device Interaction

The photograph and text-labels featured in the app would have exhibited reduced sharpness when viewed on devices having a spatial resolution of fewer than 290 PPI. According to Spencer et al. (2013), this would mean that the small anatomical structures in the learning activity would have been less easily resolved by the learners in the non-retina-display group. This would require the learners with low-resolution displays to complete more swiping and zooming of the touchscreen in order to study fine details on the image or read small text. This additional interactivity is likely to have had an effect on task-time in the same way as excessive scrolling (Molina et al., 2014; Raptis et al., 2013). In addition to contributing to physical demand, there is also the potential for impaired cognition because time spent manipulating the image is known to adversely affect the understanding and assimilation of learning materials. Molina et al. (2014) identify the time spent on such interactions as “non-useful” because the learner is not using it profitably in understanding the displayed content.

This proposition was tested post-hoc by correlating physical demand and temporal demand for each group, and although a causal relationship cannot be inferred, there was

a moderate positive correlation demonstrated between physical demand and temporal demand in the non-retina-display group, but no correlation in the retina-display group (see page 228).

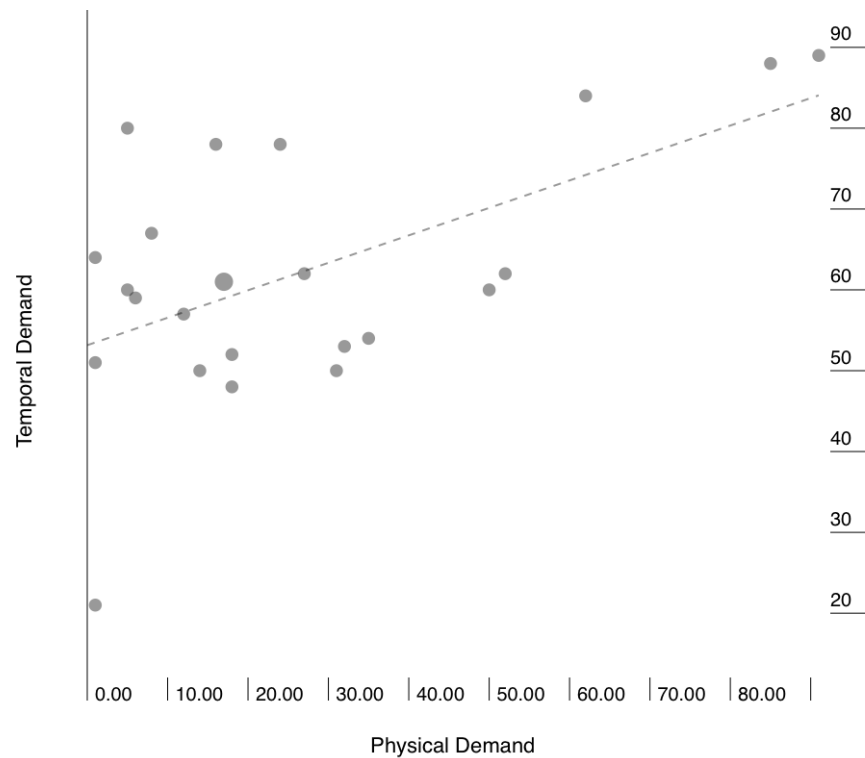


Figure 5-10: Scatterplot showing moderate correlation between physical and temporal demand in non-retina-display device learners

This correlation supports the assumption that zooming and scrolling might have increased temporal demand in device-users operating lower resolution devices.

The HCI principles of good practice indicate that the need for scrolling should be minimised in software design, especially on devices with small screens, to increase both efficiency and user-friendliness. Seong (2006) explains that this is usually facilitated by making the display area fit within the screen size. The app created for the learning activity was designed to do this by automatically adapting the photograph size to match the user display, however, this will also have reduced the resolution of the image for learners having non retina-display devices because the entire image would be represented using a

comparatively lower number of pixels. The resulting necessity to swipe and zoom is likely to have increased physical demand by a small amount as suggested by the result.

Overall Device Size

A second feasible explanation for increased physical demand in the non-retina-display group relates to device size. As stated in section 5.3, device size appeared to affect physical demand when the dimensions were too large to allow the device to be used with one hand. The data shows a strong negative correlation between screen resolution (PPI) and device size. This result appears to contradict the findings of Han and Cho (2016) who identified a strong positive correlation between screen size and PPI (Figure 5-7). However, the Han and Cho study only looked at mobile phones. When the results of my study are filtered to exclude tablets the same relationship is demonstrated, spatial resolution increases with screen size. However, my study also included tablets, and the data shows that these devices had a lower mean spatial resolution than the smartphones. Large devices have larger screens and (all other factors remaining equal) the pixels are required to cover a larger area, reducing the number of pixels per inch.

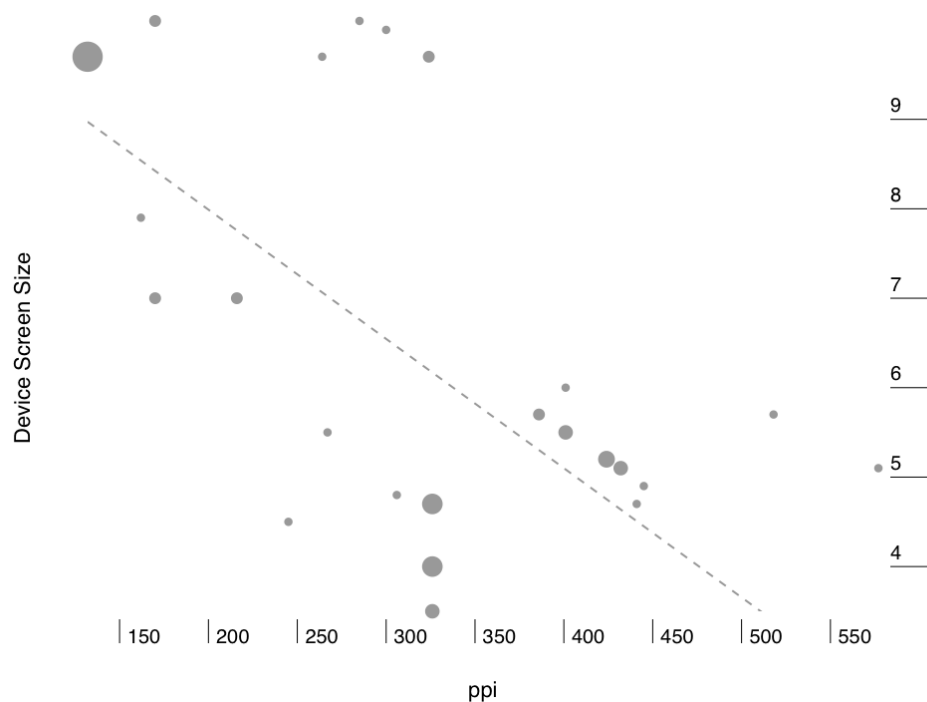


Figure 5-11: Scatterplot showing moderate negative correlation between screen size and spatial resolution

In this case, physical demand would be expected to increase because the tablet devices used in the study co-incidentally tended to have a lower screen resolution and, for reasons covered in the previous sections, these devices may be more difficult to handle increasing physical demand. When tablet devices were excluded from the data, there was no relationship between spatial resolution and physical demand amongst the smartphone-learners.

5.6.2. Evaluation of Results Relating to Correlations between Display Resolution In Pixels per Inch and the Subscales of the NASA TLX.

The scales used for task load and screen resolution are both (theoretically) continuous and are measured at the ratio level. This allowed correlation tests to be run between screen resolution and all of the dependent variables. There were three statistically significant results relating to net task load, mental demand and physical demand.

Net Task load

There was a weak negative correlation between net task load and screen resolution, showing that task load increased as screen resolution decreased.

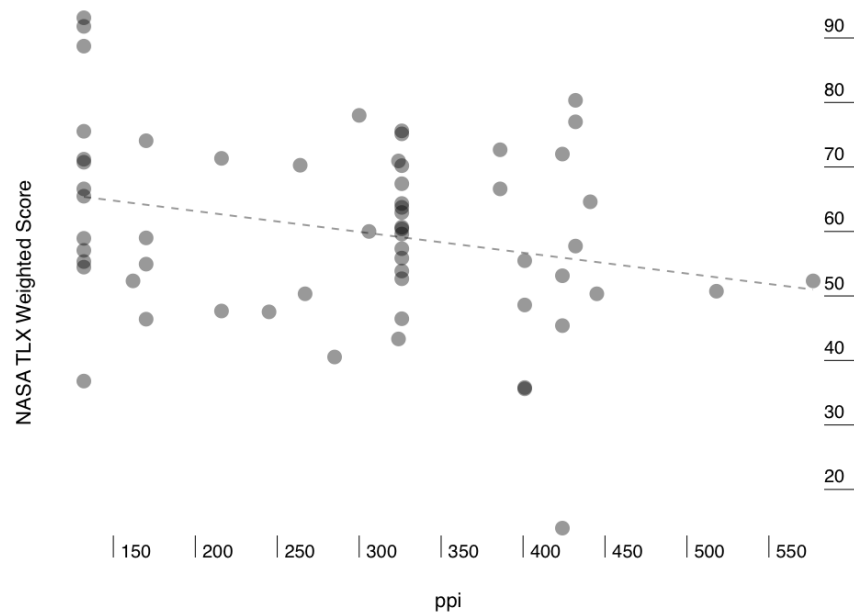


Figure 5-12: Scatterplot showing weak negative correlation between spatial resolution and net task load

The contributing factors were assessed by looking at the individual subscales of the NASA TLX. There were two statistically significant results:

Physical Demand

There was a weak negative correlation between reported physical demand and the screen resolution of the device used.

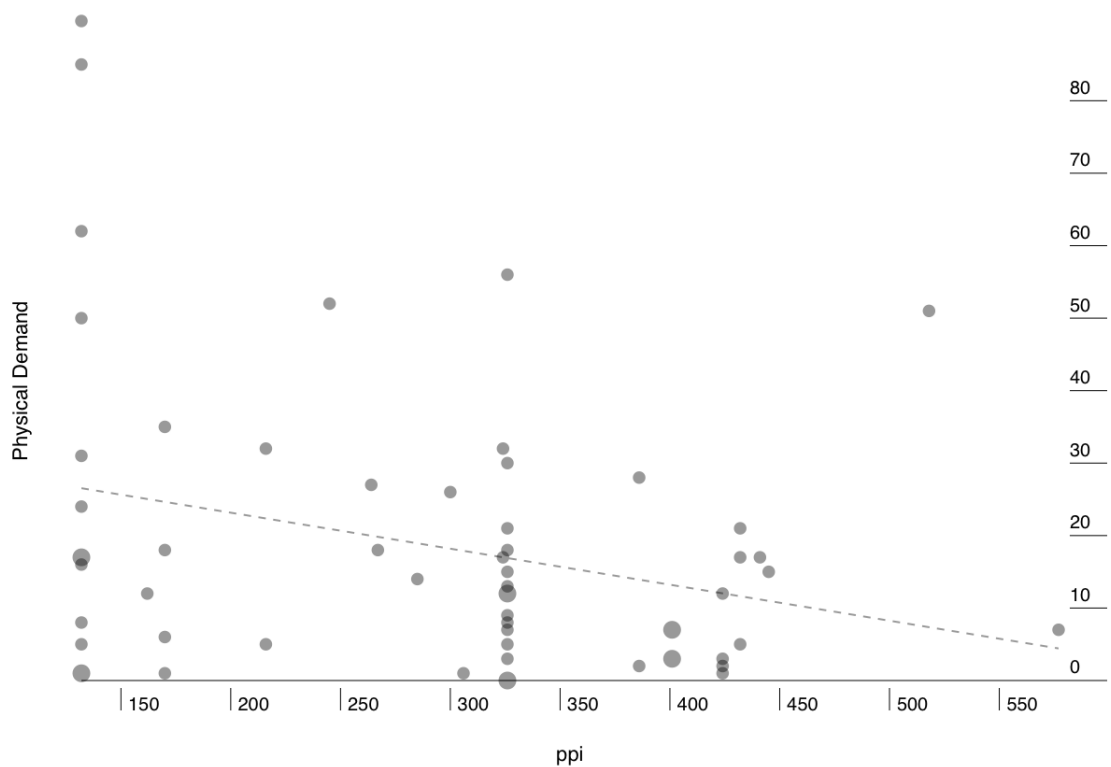


Figure 5-13: Scatterplot showing weak negative correlation between physical demand and spatial resolution

The reasons for this are likely to be the same as those articulated in the first part of section 5.6. Tablet devices used in the study tended to have a lower display resolution than the smartphones and it is likely that the physical device size was responsible for the effect rather than the screen resolution per se.

Mental Demand

There was a weak negative correlation between reported mental demand and the screen resolution of the device used ($r=-.34$, $p=.02$).

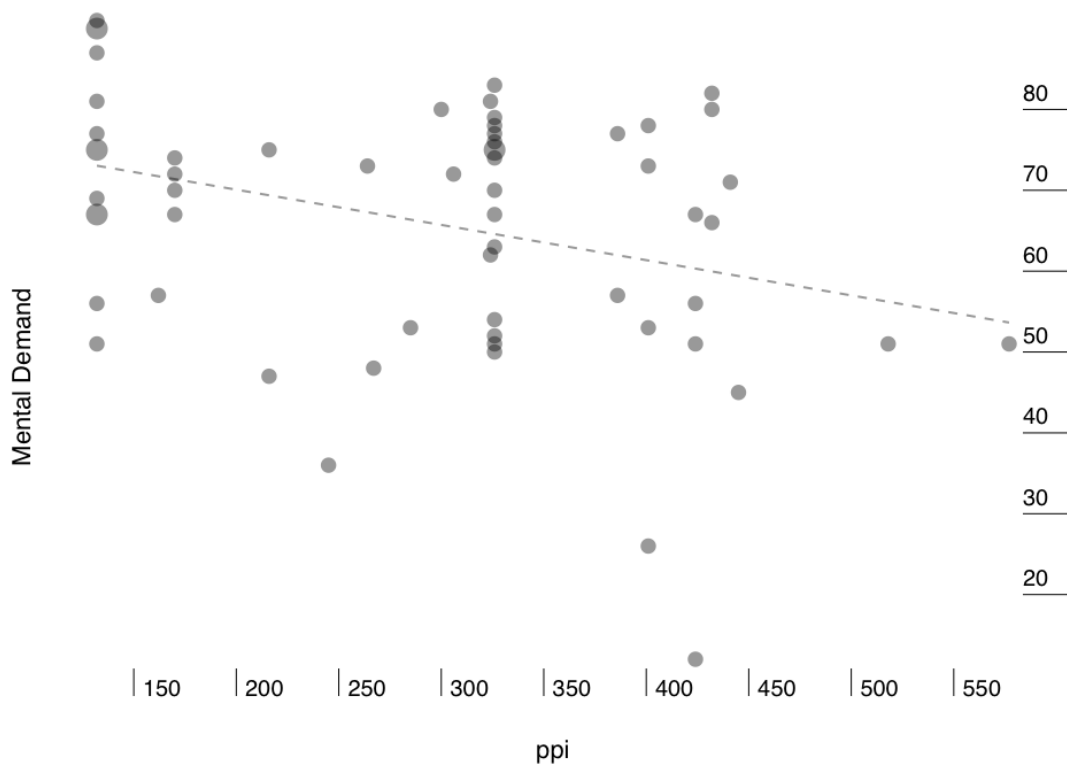


Figure 5-14: Scatterplot showing weak negative correlation between spatial resolution and mental demand

Having searched the literature for any underpinning theory or previous research into this area, there seemed to be a gap. A few authors have looked at screen-size, but there do not appear to be any studies looking at the effects of spatial resolution. In the study mentioned in section 5.5.3, Lin, Wang and Kang (2015) evaluated the differences between a tablet PC and paper-based materials used in a short-term memory test and cited decreased spatial resolution as a cause for decreasing visuospatial short-term memory performance. This is, therefore, likely to have had an impact on cognitive load and suggests that there is the need for further research into this area.

5.7. SUMMARY OF CHAPTER 05

This chapter evaluated the results in the light of the current topic literature relating to the theoretical framework used for this study, namely CLT (Sweller,1989), HCI (Longuet-Higgins,1981; Card, Newell and Moran,1983) and the Cognitive Theory of Multimedia (Mayer, 2009). In précis, the results showed that:

There was no statistically significant difference in learning task load demonstrated between mobile learners and non-interactive learners. There was a non-statistically significant higher task load reported by the non-interactive group (82% CI). This result was counter-intuitive from an HCI perspective where the use of a computer to perform the activity would be expected to increase cognitive load.

There was a statistically significant difference in learning task load demonstrated between mobile learners and non-interactive learners when data from tablet computers were not included. The non-interactive learners reported a higher mean task load score than the smartphone-learners (CI 98%). Again this result was counter-intuitive from an HCI perspective where the use of a computer to perform the activity may be expected to increase cognitive load. Three theory-based effects were presented as possible explanations, namely, the Modality Effect, the Split Attention Principle and user engagement. These effects are known to affect cognitive load in the learner and can be shown to be different between groups.

There was no statistically significant difference in pre/post-test score demonstrated between mobile learners and non-interactive learners. The mobile learners achieved a non-statistically significant higher mean pre/post-test score than the non-interactive learners (85% CI).

There was a statistically significantly higher level of mental-demand reported by the non-interactive learners. CLT (Sweller, 1989) was employed to provide a theoretical explanation for this result. Mental demand per se, is not an undesirable characteristic of a learning task, however, if mental demand exceeds the cognitive resources available to the

learner there is a corresponding decrease in performance. It was thought that low element interactivity in the m-learning task reduced extraneous cognitive load in the mobile learners.

The non-interactive learners reported a statistically significant poorer perceived task performance and this was borne out by a (non-significant) lower mean pre/post test score than the mobile-learners group. There was also a weak negative correlation between performance and pre/post-test score. CLT (Sweller 1989) was used to provide a theoretical explanation for this result. Performance increases with task load until a threshold of cognitive overload is reached. There was a significantly higher level of mental demand reported by the non-interactive group and therefore it is likely that this was the cause of poorer performance in this case.

Tablet learners reported a statistically significantly greater task load than smartphone-learners. The subscales of mental and physical demand were the main contributing factors. CLT (Sweller 1989) and the Baddeley and Hitch model of working memory (1974) were used to provide a theoretical explanation for increased mental demand, namely that the poorer spatial resolution of the tablet devices might have decreased visuospatial memory performance. HCI theory was used to suggest an explanation for increased physical demand, as tablets are too large to comfortably use with one hand (Raptis et al., 2013).

Tablet learners reported a statistically significantly greater level of mental demand than smartphone-learners. From an HCI perspective, device familiarity and preferred screen-size can affect device usability and cognitive load and the tablet-learners may not have been using devices that they were comfortable with (Koole and Ally, 2009). Another explanation lies in the fact that the spatial resolution of the tablet devices was significantly lower than the smartphones and this factor has also been shown to affect cognitive load due to decreasing the performance of the visuospatial modality in short-term memory (Lin, Wang and Kang, 2015).

Tablet learners reported a statistically significantly greater level of physical demand than smartphone-learners. This is in agreement with previous HCI research that shows device usability decreases when the size becomes too large to operate with one hand (Raptis, et al., 2013).

Tablet learners reported a statistically significantly greater level of overall effort than smartphone-learners. Effort is a combination of mental and physical demand and is explained above.

Tablet learners reported a statistically significantly greater level of task load than users of large-screen smartphones. The main contributing factor to this result was the subscale of frustration. This result was considered from an HCI perspective whereby the physical size of the device becomes too large to operate with one hand (Raptis, et al., 2013). From a CLT perspective the fact that the tablets used in this study had a lower mean spatial screen-resolution may also have caused an increase in cognitive load due to decreasing the performance of the visuospatial modality in short-term memory (Lin, Wang and Kang, 2015).

Users of devices having a high spatial resolution display (non-retina-discernible pixels) reported a lower level of physical demand than users of low spatial resolution devices (retina-discernible pixels). Two explanations were offered for this seemingly counterintuitive result.

Firstly that low screen resolution would have required the learners to interact more with device in zooming and scrolling the image to identify small structures. This additional manipulation would have decreased the time available for effective schemata formation (Molina et al., 2013; Raptis et al., 2013).

Secondly the result was only statistically significant when tablet devices were included in the data. The tablet devices used in the study had a lower spatial resolution than the smartphones and due to their larger physical size would have increased physical demand (Raptis, et al., 2013).

6. CHAPTER 06 - CONCLUSION

This chapter is intended to provide a brief overview of the intentions and goals of the research. There is a critical reflection on the research process, considering reliability, validity and also an explanation about how methodological issues and practical difficulties were addressed during the study. Factual conclusions are presented relating to the research hypotheses and how the research questions were answered to provide justifiable conclusions. Secondary conclusions are also analysed. Generalisability is discussed, and relevance of the findings to the wider field of m-learning and application design are identified. A modest contribution to knowledge is proffered in the light of these conclusions. An agenda for potential further research is presented based on the findings and conclusions of this study, and also where unexpected secondary findings suggested an opportunity for research that was not within the boundaries of this investigation.

As a professional doctorate, the final section of the conclusion is a reflective, critical evaluation of the contribution which this research makes in the context of professional practice. This includes a critical account and analysis of personal intellectual and professional development during the research process.

6.1. RESEARCH PURPOSE AND INTENTIONS

The purpose of this study was to empirically investigate the use of mobile touch-screen devices as tools for delivering learning materials. The particular focus was on whether smartphones or tablet computers had any effect on learning task load or offered any measurable benefit to the achievement of successful learning outcomes when compared to traditional textbook learning. On first thought, this seems like an obvious question that might have already been investigated. However, when this study was commenced in 2012, tablet computers were still relatively new to the market. Apple's iPad had been released just two years previously, and research into touch screen devices was still limited. Kukulska-Hulme, et al. (2010) identified that this mode of learning was still a novelty for many educators and that as a concept that focuses on "learning rather than teaching" there were challenges to educators in understanding how this technology

might address the needs and abilities of learners. A scoping review of the limited literature available showed that there was an emerging theory of m-learning, initially defined by Sharples, Taylor and Vavoula (2007), that was critically defined by the fact that learners were on the move, and not necessarily inside a classroom. The concept of situated learning and the affordances offered by mobile devices in this regard was the focus for many of the early studies in the field (Sølvberg and Rismark, 2012; Terras and Ramsey, 2012; Wang and Shen, 2012; Martin and Ertzberger, 2013; Chu, 2014). Kearney, et al. (2012) recognised the ubiquity of mobile devices as providing an ideal tool to mediate socially-constructed learning in the tradition of Vygotsky (1986). Other early research looked at topics such as user engagement (Pachler, Bachmair and Cook, 2010; Deegan and Rothwell, 2010; Manuguerra and Petocz, 2011; Gikas and Grant, 2013) and from a pedagogical viewpoint there were studies that explored how educational theory might be adapted for the new digital media (Laurillard 2007; Pachler, Bachmair, and Cook 2009; Sharples, Taylor, and Vavoula 2007).

However, there seemed to be a gap in the literature regarding how mobile devices might impact on the cognitive processing of the user. This was surprising because cognition is a prerequisite for learning and HCI theories acknowledge that mobile devices, as computers, might be expected to present cognitive barriers (lack of user-friendliness) that are not found in traditional learning materials such as books (Card, Newell and Moran, 1983). Following the introduction of the iPad into the mass market in 2010 and the subsequent proliferation of touch-screen computing devices, a need for research into the area of cognition in m-learning was identified. Hollender et al. (2010) and Schmidt-Weigand and Scheiter (2011) posited that the increasing complexity of virtual learning environments was likely to have an impact on cognitive load in the learner. Terras and Ramsey (2012) also recognised this need for further research and identified five specific challenges that are unique to m-learning and which affect the learner in ways that educators may not have previously needed to consider. Four of these challenges related to cognition (*the context-dependent nature of memory, the finite nature of human cognitive resources, distributed cognition and situated learning and metacognition being essential for m-learning*) and recognised that the multimedia presentations that are

required to provide a “rich” learning experience might also adversely affect cognitive load. For this study, it was therefore decided to perform an experiment to investigate cognitive load in the mobile learner. A professional doctorate must, by definition, relate to the professional practice of the candidate. From a professional viewpoint, there was also a sound rationale for looking at cognition relating to m-learning. The researcher has a background in neuroscience and a grounding in cognition from a neuroscientific perspective and has also been involved in the development of e-learning materials for approximately 18 years. These materials include applications for smartphones and tablets and electronic editions of three popular textbooks available as Kindle and iBook editions. As an app developer all of the learning materials created by the researcher are in a form that can be accessed on mobile platforms such as tablets and smartphones. As a textbook author, the researcher was interested to note that much of the research that had been conducted on mobile devices tended to compare smartphones and tablets with other digital media such as personal computers, or used PC simulations (Findlater and McGrenere, 2008; Kim and Kim, 2012; Molina et al., 2014). It was thought that this was a somewhat invalid comparison because the advantages of m-learning, as identified in the topic literature, centre around the portability and ubiquity of the devices used (Sølvberg and Rismark, 2012; Wang and Shen, 2012; Martin and Ertzberger, 2013; Chu, 2014). These advantages are not applicable to personal computers, or perhaps even so-called “notebook” laptops – as these machines are not as ubiquitous as smartphones are larger than smartphones and are not pocketable (Shirer, 2015). Historically, the tablet PC was never intended to replace the computer, like the Kindle tablet and iBook apps of today, the aim of the device was to replace the book, while offering a new level of user-interactivity via multimedia (Kay, 1972).

The overall intention of this study was, therefore, to compare cognitive task load between two groups of learners, an interactive group of mobile-learners using touchscreen devices and a non-interactive group using an appropriate comparator - a labelled photograph as might be included in a traditional textbook. The full list of aims and objectives can be found on page 2, and the full list of hypotheses in section 3.6. The empirical nature of the data dictated the use of a post-positivist research approach whereby quantitative methodology was employed. The study employed a cross-sectional,

two-armed controlled trial designed to identify, measure and compare differences in levels of self-reported task load between two parallel, balanced groups of learners during a learning task. The learning activity in question required the participants to learn the names of the foramina (openings) found in the base of the human skull and memorise the names of the blood vessels and cranial nerves that pass through these spaces. This represents a typical learning task found in the study of anatomy and has a moderately high degree of element interactivity that might be expected to challenge the cognitive resources of the learners.

In addition to this secondary data was collected relating to the screen size of the devices used and the spatial resolution of the device displays. These variables were also tested for any effects on task load and across the subscales of the NASA Task Load Index.

6.2. RESEARCH BOUNDARIES

6.2.1. Learning Theory

The research had well-defined boundaries inasmuch as the learning activity represented a rote-learning approach typically found in anatomy lessons. Mobile devices have also been found to offer advantages in constructivist learning, but this learning theory was not within the bounds of this project. The data collected looked at task load and used theories relating to human cognition. It is acknowledged that there are other learning theories and that this only represents a cognitivist approach to learning.

6.2.2. Generalisability

The participants enrolled in the trial were intended to represent distance learners in the field of medical and healthcare education. The sample used a demographic that was of a representative age group (36 years) and gender-mix (female 71%, male 29%) for this group of learners and it is not proposed that the results obtained in this study would necessarily be generalisable to learners in other age groups or disciplines. There were no statistically significant differences demonstrated between participants grouped by gender or by age.

6.2.3. Methodology

The data collected for the study was entirely quantitative. It may have been useful to employ a mixed-methods approach to help identify specific reasons for increases in task load for the subscales of the NASA TLX. Practical considerations made this unfeasible due to timing and financial constraints. It would have been necessary to interview participants in addition to the quantitative data collection tasks. It is likely that this would have increased data collection from 40 minutes to over two hours and would have placed an unacceptable burden on the participants. It would also have required hiring the conference facilities (where data collection took place) for a further two hours each session, and this was beyond the means of the researcher. It is acknowledged that a mixed-methods approach might have offered further insights into the reasons behind

some of the findings of the study, particularly the differences between smartphones and tablets. This may present an avenue for further research.

6.3. CONCLUSIONS

6.3.1. Answer to The Research Question

The primary research question was:

“Is there a statistically significant difference in the level of task load (cognitive load) experienced by a learner when undertaking a multimedia interactive learning activity delivered by a mobile touch-screen computing device compared to that experienced by a learner undertaking an equivalent non-interactive learning activity designed to teach the same factual information?”

From the results, the justifiable answer to the primary research question is that there was no statistically significant difference in task load between the mobile-learners and the non-interactive learners. However, when tablet computers were excluded from the data analysis, there was a statistically significant difference between the groups. The smartphone-learners reported a lower mean task load than the non-interactive learners and the difference in mental demand became more pronounced. The smartphone-learners reported a significantly lower mental demand than the non-interactive learners. Although a two-tailed test was used, and there were no formal assumptions that there would be any direction to the result, the data showed that the net task load was lower for the mobile-learners even when tablets were included. This result was not statistically significant at the 95% confidence level, but the direction of the result suggests that, if anything, the learning task placed a greater load on the non-interactive learners. When the tablet-learners were excluded from the data analysis, the result became statistically significant at a confidence level of 95% ($p=.040$). Although the effect size was small, this direction in the result was unexpected, because the control group were only required to study a labelled photograph, whereas the experimental group were required to interact

with a multimedia presentation on a mobile device. Attempts had been made to control any other variables that might have affected task load – such as distraction, and time pressure. As a computer, the use of a mobile device might be expected to contribute to cognitive load (Card, Newell and Moran, 1983) and the use of multimedia in learning materials is also known to affect cognitive load (Mayer, 2009). A performance outcome measures (pre/post-test) was used to support any findings from the NASA TLX. The results of this test showed no statistically significant difference between groups and support the factual conclusion that the use of a mobile device does not have any measurable effect on task load or learning outcomes in comparison to studying a traditional textbook-style labelled photograph. The data analysis showed that mental demand was significantly lower in the mobile learners' group and the mobile-learners also reported a higher level of perceived performance.

[Recommendations Relating to the Primary Research Question](#)

For e-learning developers, the importance of this result in the context of the study is twofold:

Firstly it justifies the potentially long development-time for e-learning materials and their associated multimedia content. There is no inherent disadvantage to the learner when using a mobile device instead of a book diagram, and the use of a smartphone reduces task load compared to learning from a book diagram.

Secondly, it highlights the importance of good practice in the design of multimedia content. The net cognitive load imposed on the learner during a m-learning activity is the sum of many effects related to the physical attributes of the device used, the ease of use of the interface and the complexity of the learning task (Hollender, et al., 2010). Learning materials that follow the principles of the Cognitive Theory of Multimedia (Mayer, 2009) are likely to reduce cognitive load in the learner and counterbalance any other causes of extraneous cognitive load induced by the mobile device.

From a textbook author's viewpoint, the results suggest that not only are smartphones a suitable medium for delivering written content, there is also value in making electronic publications more interactive. In the previous discussion chapter it was posited that the use of audio is likely to have reduced cognitive load due to the advantage of dual coding

(Paivio, 1990) and although this should be used with care to avoid redundancy effects, it may provide an effective augmentation to text publications. Further research into this area may be useful to specifically identify how the inclusion of multimedia may reduce cognitive load for smartphone-learners accessing electronic versions of textbooks.

6.3.2. Secondary Conclusions

Smartphones vs. Tablets

There was a statistically significant greater task load reported by the tablet-learners compared to the smartphone-learners. This result was confirmed in the test that compared task load across all of the screen sizes used by the participants. The main contributing factors were both mental and physical demand. The learning activity was identical for both groups, so the reason for an increase in mental demand is likely to relate to the device characteristics rather than the learning materials or software design. It was noted that the smartphones used in the study had a substantially higher spatial resolution than the tablets. The mean ability to demonstrate small detail was 1.8 times higher in the smartphones. Effective learning requires optimisation of working memory resources, and according to Baddeley and Hitch (1974), visual input relates to two important channels of working memory. The phonological loop deals with written materials (in addition to spoken words) such as diagram labels. The visuospatial sketchpad processes information from visual imagery. It is therefore posited that, in agreement with Lin, Wang and Kang (2015) the reduced spatial resolution of the tablet devices is likely to have caused a corresponding decrease in visuospatial memory performance.

Previous research by Raptis et al. (2013) highlighted the fact that increased physical demand associated with tablet computers reflects the fact that they are of a size whereby the device can no longer be operated using one hand. This result may also, therefore, support findings by Periera et al. (2015), who demonstrated that large tablets reduced overall usability and productivity, increased user-fatigue and that smaller devices could be held comfortably for twice as long as large tablets.

Recommendations Relating to Device type

There are three recommendations that can be made in the light of this finding. These pertain to interface design, material design and the choice of hardware when providing mobile devices to students.

Regarding interface design, e-learning materials must be responsive to screen size and whether delivered via mobile app or browser should adapt to the size and resolution of smartphone screens. Responsive design ensures that content such as text and graphical imagery are optimised for the screen size and spatial resolution of the device being used. A recent working group paper by ECAR (Educause Center for Analysis and Research) highlighted the need for responsive web design in higher education (Bollens, et al., 2014). This finding is relevant to material design in that images that are presented in the learning materials should take advantage of the native resolution of the device display. Modern devices, particularly smartphones, are capable of high spatial resolution and there would be no advantage in presenting an image with a spatial resolution of 300 dots-per-inch on a device that is capable of screening a resolution of 370 pixels-per-inch.

Regarding hardware choice, some universities supply touchscreen devices to students to use in their studies. If smartphones reduce learning task load compared to tablets it may affect the choice of device to be supplied, or indeed the decision as to whether tablet devices should be supplied at all. Although there is controversy, in the literature review (Chapter 2) there were studies that seemed to suggest that smaller screens were better suited to learning than tablets (Martin and Ertzberger, 2013), that students did not readily accept tablets as learning tools (Keller, 2011) and that touchscreen iPods (which have a small screen size) were motivational, empowering and were responsible for a marked improvement in learning outcomes (Clark and Luckin, 2012). Furthermore, Raptis et al. (2013) also noted that learners preferred a screen-size of 4.0" - 5.0" and that a higher usability-rating was reported by those who used their own device.

When tablet-learners were compared to non-interactive learners in this study, there were no statistically significant differences demonstrated in any of the sub-scales of the NASA TLX with the exception of physical demand which was higher in the tablet-learners as might be expected. These findings must be taken into consideration when taking the decision to purchase high-cost tablet computers that may not offer any advantage over the students own smartphone or traditional learning materials such as books. It is

recognised that this study only considered one type of learning activity, however, and it may be an agenda for further research to investigate whether tablets offer advantages over other modes of delivery when undertaking different types of learning activity.

[Spatial Resolution of Device Display](#)

The major manufacturers of touchscreen devices now offer non-retina-discernible pixel densities on their device displays. This feature provides a high spatial resolution even at a short viewing distance of 30 cm. There were no differences reported between learners when grouped into “retina-display” vs. “non-retina-display” with the exception of Physical Demand. The “non-retina-display” group (discernible pixels) reported a greater physical demand than the “retina-display” group (non-discernible pixels). It is thought that this finding is coincidental with the fact that the tablets used in the study tended to have a lower screen resolution than the smartphones, and the result is due to the physical attributes of the tablet rather than the screen resolution being used.

When the pixels-per-inch values were analysed as a continuous scale, there were two statistically significant negative correlations demonstrated between spatial resolution and the NASA TLX sub-scales of Mental Demand and Physical Demand. These were notably the same dimensions that were affected when comparing tablets and smartphones. With this in mind, the results from the tablet devices were removed from the data, when comparing smartphones only, there was no correlation found at any confidence level between the pixel density of the device and any subscale of the NASA TLX. This supports the previous finding that any differences are likely to be due to the physical attributes of tablet computers rather than the coincidentally lower spatial resolution of the tablets used.

[Recommendations Relating to Spatial Resolution of Device Display](#)

This result appears to support the previous findings, that smartphones offered an advantage over tablets in reducing physical and mental demand. The result should be treated with caution because more recently-released tablet devices tend to feature higher resolution displays than the models used in this study. This improved level of detail might have had an effect on mental demand for reasons covered on pages 262 and 275. However, taken in context with previous findings from the topic literature (Keller, 2011;

Clark and Luckin, 2012; Martin and Ertzberger, 2013; Raptis et al., 2013) this finding may support the proposition that smartphones are a better choice than tablets for delivering learning materials and offer a potential empirically-determined explanation.

6.4. ORIGINAL CONTRIBUTION TO KNOWLEDGE

6.4.1. Theoretical Contributions

The literature review presented in Chapter 02 indicated that there had, to date, been no studies conducted to evaluate the cognitive load associated with mobile devices when used as learning tools. Much of the existing research compared mobile devices to personal computers or laptops. These comparisons ignored the primary benefit of mobile devices as portable, ubiquitous content-delivery systems and failed to consider books as a more appropriate comparator. Previous studies also tended to use small sample sizes, convenience sampling and recruited participants that were not of a representative age-group for mobile learners. Previously conducted studies identified the need for further research into screen size, cognitive outcomes relating to learning from mobile apps and screen resolution when learning anatomy.

From the results provided in the previous chapter, it is proposed that the following modest contributions to knowledge have been made:

1. The study provides empirical evidence that the use of a mobile device does not have any measurable adverse effect on learning task load or learning outcomes when compared to a non-interactive mode of learning, namely, studying a textbook diagram.
2. The study provides empirical evidence, with a small effect size, that smartphone-learners experience a lower mean task load during a learning activity when compared to a non-interactive mode of learning, namely, studying a textbook diagram.
3. The study provides empirical evidence, with a medium effect size, that mobile-learners experience a lower mean level of mental demand during a learning activity when compared to a non-interactive mode of learning, namely, studying a textbook diagram.
4. The study provides empirical evidence, with a small effect size, that mobile learners report a higher mean level of task-performance during a learning activity

when compared to a non-interactive mode of learning, namely, studying a textbook diagram.

5. The study provides empirical evidence to show that smartphones offer advantages over tablet computers in a learning activity that requires rote learning.
Smartphones induce lower levels of net task load, mental and physical demand on the learner, and reduce the level of overall effort expended on a learning task when compared to tablet computers.
6. The study provides empirical evidence to show that increasing spatial resolution of the device display negatively correlates with levels of mental demand, physical demand and net task load.

6.4.2. Methodological Contributions

1. The study has shown that an on-line tool can be used to deliver a learning activity and assess cognitive load and performance outcomes and analyse data in a single application. However, it has also shown that the use of such a tool may not harvest data as quickly or accurately as face to face data collection.
2. The study used authentic devices rather than simulations when investigating spatial resolution and used a data collection tool that was appropriate for assessing multi-dimensional task load associated with Human-Computer Interaction.
3. Unlike previous similar studies, the data was collected from an appropriate demographic for distance learners from the health sector, namely post graduate learners with a mean age of 36 years and an approximately representative gender ratio of 6:3 (female: male).

6.4.3. Contributions to Professional Practice

These findings also contribute to professional practice in that they permit experimentally-justified choices to be made when creating mobile apps for the delivery of learning materials. Firstly, that smartphone m-learning offers advantages over traditional textbook learning in the cognitive domain. Secondly, that m-learning apps should be optimised for smartphones and should take advantage of the high spatial resolution of modern device displays. These contributions to professional practice are critically reflected upon in section 6.7.2.

6.5. CRITIQUE OF THE RESEARCH

Every effort was made to ensure that the research was conducted according to best-practice. However, there were some issues encountered during the research process, and also some extraneous variables encountered that could not be fully controlled. Some potentially confounding variables were only recognised after data collection had been completed. There is a small possibility that some of these issues might have affected the validity and generalisability of the research. The following critical evaluation explains how the issues were resolved.

6.5.1. Validity

Content Validity

Content validity (whereby a measure incorporates all of the necessary elements of a particular construct under investigation) was considered in choosing a data collection tool for the study. The multi-dimensional nature of the NASA TLX was thought to offer a greater degree of content validity than, for example, a post-task questionnaire that only sought to assess mental load. This is because other factors such as physical and temporal demand may also contribute to overall task load.

Construct (Factor) Validity

In the subjective measurement of task load, the NASA TLX is considered to have the highest factor validity (i.e. the highest correlation with the factor it was intended to measure) (Noyes, Garland and Roberts, 2004). The originally proposed choice of data collection was to use functional MRI. However, this was not considered to be ethically justified. Furthermore, a recent study by Eklund, Nichols and Knutsson, (2016) has shown that the most commonly used software packages for analysing fMRI data can result in false positive rates of up to 70%. This figure had previously been assumed to be as low as 5% and has called into question the validity of many previous studies. For this reason, the use of NASA TLX over fMRI was a serendipitous decision.

Data Validity

It was noted during the data collection activities that some participants were not performing the tasks correctly. In some cases, participants returned to the pre-test after having undertaken the learning activity or attempted to write down the answers during the learning activity. These cases were picked up by invigilation, and the papers were excluded from the data analysis. Data collection instructions were modified to prevent incorrect data entry for further data collection sessions. Participants were asked to draw a line through any answer boxes that could not be completed in the pre-test preventing them from filling them in retrospectively.

Statistical Validity

To ensure confidence in the methods used, a medical statistician with expertise in randomised controlled trials was consulted to provide retrospective advice on the data analysis undertaken and to highlight any potential weakness or limitations relating to the approach used. The statistician confirmed that the appropriate tests had been conducted, the necessary post-hoc data analyses had been undertaken correctly and that all of the assumptions relating to the data had been met. The decision to present the data for both parametric and non-parametric tests was considered to be justified because only data that was obtained using the appropriate test for each sample was used in the results. Furthermore, Norman (2010) explains that the widely-held belief that parametric tests

must be based on an assumption of normality is not strictly correct. For t-tests and ANOVA, it is the normality of the *distribution of the means* that is assumed, not the data. Norman goes on to say that there is a wealth of empirical data to show that parametric tests examining the difference between means do not require an assumption of normality unless the sample size is less than $n=6$. Presenting the parametric test results, therefore, served to reinforce the non-parametric test results as they were almost entirely in agreement.

For the secondary data analysis, comparisons were made between devices and between screen sizes. For the between-device tests, the mobile learners were subdivided into two subgroups, namely smartphone users and tablet users. For the screen-size tests, the mobile learners were subdivided into four groups representing small, medium, large and very large screens. Subdividing a group into a number of smaller subgroups introduces an increased potential risk of type I (α) and type II (β) errors, namely false positives, and false negatives. In turn, this tends to reduce the statistical power of a test because power is stated as $1-\beta$ (Hawkins, 2009).

For the comparisons between tablet users and smartphone users the group sizes were still large enough ($n=24$, $n=36$) to obtain statistically significant results, however, the effect size was small. When an effect size is small, it is usually desirable to employ larger sample sizes where possible. For this reason, the statistical power was lower than for the main data analysis. A post hoc test revealed that the statistical power had fallen to 0.7. For this reason, it was articulated in chapter 04 that these results should be viewed with caution. In the opinion of the statistician, these results are fit for presentation in the thesis because, as new research data, they are still of value - particularly as a baseline for further studies. For future research, the post hoc analysis indicated that a statistical power of 80% would be restored by using a sample size of 152 participants ($n=76$ in each group). Notwithstanding this post-hoc finding, a fair degree of confidence can be attached to these between-device results because the screen-size tests described below support some of the findings with a large effect size and higher statistical power.

For the screen size tests, the mobile learners were subdivided into four groups ($n=10$, $n=13$, $n=17$, $n=20$). This lead to multiple comparisons using ANOVA and Kruskal-Wallace tests for normal and non-normal data respectively. Multiple comparisons were not made

with the control group because these learners did not use devices. Variables such as screen size and resolution were therefore not meaningfully comparable. By design, tests involving multiple comparisons test multiple hypotheses which increases the risk of finding a rare event (Abdi, 2007). The statistician confirmed that this effect (known as inflation of the alpha level) was reduced in my study by my having applied a post-hoc Bonferroni correction to the data. Despite the smaller group sizes, the statistically significant results from the screen-size comparisons showed large effect sizes for net task load ($d=1.131$) and frustration ($d=1.186$). The post-hoc tests for statistical power for these comparisons indicated values of 0.79 and 0.84 respectively. This is in agreement with an a-priori, sample-size calculation using G*power software (Faul, et al., 2009) that indicated a required total sample size of $n=60$ (as used). These results can therefore be considered robust.

For future studies, the statistical power could be improved for the secondary between-device tests by increasing the total sample size of mobile learners to $n=152$. In view of Norman's (2010) explanation relating to parametric and non-parametric testing, it may be considered acceptable to use parametric tests for non-normal data where the sample size is large enough.

Ecological Validity

The original on-line data collection tool posed an intrinsic threat to ecological validity in that the labelled photograph was intended to be presented as a static, non-interactive image on a computer monitor. This would not have been an optimal comparator for a textbook photograph because it is likely to have been presented at different display sizes and spatial resolutions according to the user's monitor specifications. This shortcoming was addressed in the revised data collection method by using an identically-sized labelled photograph, having the same spatial resolution for every data collection session. There was a remaining concern relating to ecological validity that resulted from the need to control the test conditions, namely that the mobile learners all undertook the activity in a classroom setting. While it is recognised that mobile devices have a place in classroom teaching, it is an accepted benefit of m-learning that the devices are used in locations other than the classroom. Had the m-learners been allowed to use the devices in the field

there might have been other variables such as distraction that might have affected the overall result. Whether it is advantageous to control such variables is debateable. If one takes the view that distractors are an inherent part of m-learning, such variables could have been included in the learning activity. The view taken in this study was that it was more appropriate to control variables that did not strictly relate to the question being asked. This study was interested in finding any differences due to hardware and software design rather than environmental distractors. Furthermore, environmental distractors would have applied to the control group too, as they represented distance-learners using textbooks which can also be accessed in locations remote from the classroom. In summary, ecological factors were controlled equally for both groups, and therefore ecological validity was maintained.

External Validity (Generalisability)

Efforts were made to ensure that the participants were representative of the population under investigation. Statistical power calculations were used to determine the group sizes, and appropriate tests were used in the data analysis. However, the group sizes used for the between-device tests were not as large as the group sizes used for the main research question. This was because the data relating to devices could only be obtained from the participants in the experiment group. 36 participants used smartphones, 24 used tablets. The imbalance being due to random allocation, data having to be discarded due to invalidation and failure of participants to declare the type of tablet used. Failure to identify the device model precluded the data from being used in the between-device tests because the screen size and spatial resolution could not be determined. This, in turn, reduced the m-learners group size to 60 participants for device comparison tests. For this reason, any results relating to screen size and spatial resolution in the between-devices tests may not be as generalizable to the larger population as the results relating to the main research question which was between equal-sized groups having a larger more appropriate sample size. Notwithstanding, the sample sizes used for the between-devices tests are equivalent to many other previous studies in the field (Raptis, et al., 2013; Chu, 2014; Molina, et al., 2014; Lin, Wang and Kang, 2015) and were found to be acceptable in a post hoc test using G*power software.

The style of learning used in the activity related to memorisation of facts relating to anatomy and was a cognitive activity rather than a constructivist learning activity. The results of the between groups tests may not, therefore, be generalizable to other modes of learning.

6.5.2. Reliability

Inter-Rater Reliability

Inter-rater reliability was not an issue because the data was empirical and did not rely on interpretation. Data analysis was carried out by a single researcher and accuracy was ensured by repeated re-checking of the NASA TLX data and re-marking of the pre/post-tests until no further errors were discovered.

Measurement Instrument Reliability

The full rationale for using The NASA TLX is covered on page 119. Previous research in the field of cognitive load has found this tool to be valid reliable and unobtrusive and have the highest factor validity and test-retest reliability. (Paas and van Merriënboer, 1993; Kaluga, Chandler and Sweller, 1998; Noyes, Garland and Roberts, 2004; Haapalainen, Kim and Dey, 2010).

6.5.3. Bias

Attrition/Recruitment Bias

The original research design intended to use an on-line data collection tool. This was a browser-based application that the participant could access via a passcode. The entire on-line process was designed by the researcher to take the participant through the information sheet, consent form, pre-test, learning activity, NASA TLX questionnaire and post-test collecting the data and storing it in an encrypted on-line database for retrieval and analysis. There is a detailed explanation of the on-line data collection tool provided in Appendix A.

It was thought that the on-line method would have allowed data collection from a large number of participants in a very short time. However, this was not the case. Over 2,400 potential participants were invited to take part, 196 agreed to participate, but after three months of data collection, only 14 participants had undertaken the on-line task. It was apparent that there would not have been enough time to complete data collection within the 12 months available. It was thought that the design of the on-line collection tool might have led to self-selection bias, in that individuals with an interest in computing and mobile devices may have been more motivated to participate. Furthermore, it appeared that some participants had started the activities and dropped out of the data-collection before the end of the on-line process. The issue was resolved by undertaking data collection in a live setting where participants could be recruited from attendees of a series of international study days being organised by the researcher. This strategy allowed enough participants to be recruited to achieve the group sizes required (total $n=130$). Although consecutive sampling is not considered to be as robust as true random-sampling, it was thought that the benefit of using the large group sizes achieved would outweigh any potential sampling bias.

Channelling bias

In order to randomly allocate participants to either the control group or experimental group, it was necessary for each participant to bring a smartphone, a tablet and a set of earbuds or headphones to the data collection sessions. Participants were made aware of this several weeks in advance, and it was explained that they did not need to bring a device if they did not wish to participate. It became clear at the first data-collection session that some participants had forgotten to bring a device, or had brought a smartphone but did not own a tablet. Some participants brought the devices but forgot to bring headphones. One participant who was randomly allocated to the tablet-user group had a smartphone but no tablet. This posed a problem because not only would it prevent willing participants from taking part; it would have made the sampling non-consecutive, and group allocation would not have been random. This might have introduced channelling bias.

To solve this issue in further sessions, it was necessary for the researcher to purchase a number of spare tablet devices and 20 sets of headphones. This allowed randomisation of

the groups as the spare devices could be offered to participants who were allocated to the experimental group who had not brought the correct device or had forgotten to bring headphones. An unforeseen trade-off was that some participants provided with a tablet failed to identify the make of the device on the data collection form.

Measurement Bias

The same data collection tools were used for both groups. Trusted and well-respected tools were employed. Repeated marking of the pre/post-tests was undertaken until no further errors encountered (4x marked).

Procedural Bias

In the original on-line data collection method it would not have been possible to monitor the test conditions under which the participants were undertaking the learning activity. By changing the data collection method, it was possible to control the data collection location, reducing the risk of procedural bias caused by on-line participants taking the tests under non-identical conditions. Activities were undertaken simultaneously by both groups, under identical conditions and using the same timing.

Reporting Bias

The risk of reporting bias was reduced by ensuring that standardised statistical tests were employed and all results were reported. All interpretations of the data were related to findings of previous studies and discrepancies and disagreements were reported in addition to findings that were in agreement with current thinking in the field.

Selection Bias

The originally-proposed on-line data collection tool may have introduced self-selection bias due to perceived computer self-efficacy (Compeau and Higgins, 1995) whereby only participants who are comfortable with mobile devices might have decided to undertake the activities. This method was discontinued and revised.

Group allocation was randomised, however, consecutive sampling was employed rather than truly random sampling. Because every available participant was included over a 12

month period (i.e. the entire pool of available participants from which a random sample could have been drawn) it was possible to retrospectively pseudo-randomise the samples by extracting every third participant from each group (using a random number generator) and running the same tests on the randomised samples. The disadvantage of randomisation was a reduction in group size ($n=21$ vs. $n=65$) but the results showed no statistically significant difference to those obtained from the larger consecutively-sampled groups. Allocation to groups and devices was randomly implemented using a random number generator.

Subjectivity Bias

The NASA TLX is a subjective measure of task load. However, the same test was used for both groups so any subjectivity bias would be expected to be the same for both groups. Objective measurement of cognitive load (pre/post-testing) was used to triangulate with the self-reported subjective task load data. This did not support the subjective findings for all of the dimensions included in the test, notably as a performance-outcomes measure of task load when comparing smartphone-learners and non-interactive learners. This was thought to be due to the difficulty of the pre/post-test subject matter and the resulting fact that data was non-normally distributed in the non-interactive group. If a similar study was to be undertaken in the future, it is recommended that an easier test topic should be employed.

In summary, the data-collection activities posed many challenges. Many of these issues were anticipated and controlled for. Unexpected problems were identified early in the data-collection process and were ameliorated. Potential causes of bias and threats to validity were therefore minimised. It is therefore expected that the results from this study are meaningful and are likely to be generalisable to distance-learners undertaking this type of learning activity.

6.6. AGENDA FOR FURTHER RESEARCH

Some of the study findings were unexpected, and at least one potentially confounding variable was identified. Some of these findings suggested possible avenues for further research; these are identified below:

Correlations between cognitive load and screen size

As noted on page 264, there were correlations made between device screen size and other variables relating to task load. A strong correlation does not indicate causality, and it was suggested that any effects relating to screen size could be caused coincidentally by other factors that also have a relationship with screen size. There has been a manufacturing trend of increasing screen size in smartphones over the last seven years, and other factors have also changed over this time scale, such as processing power, improvements in random access memory (RAM) and improved spatial resolution of the device display. Further research could be conducted to investigate the effects that these improvements might have on cognition, and device usability as a learning tool.

Temporal Demand

The data shows that mobile-learners reported a higher level of temporal demand than the non-interactive learners (i.e. felt that they were under more time pressure) than the non-interactive learners. As stated in section 5.2.8, Tsao, et al. (1983) demonstrated that if a group of participants is required to undertake an activity or stimulus that increases mental load, the duration of a secondary event is underestimated. This effect has been confirmed in later experiments (Brown, 1997). In this case, the non-interactive learning task (that increased mental demand) should have been estimated to pass more quickly. However, this was not the case, the mobile-learners reported a lower mental demand than the non-interactive learners, but indicated that they felt under more time pressure. Although this result was not statistically significant the confidence level suggests that there is an 88% likelihood that the effect was not due to chance and further research could be considered to confirm this effect and provide an explanation.

[Tablet Computers vs. Smartphones](#)

There were some unexpected results demonstrated when comparing tablet computers and smartphones. For example, overall task load was reduced as screen size increased from small smartphones to larger screen models but then significantly increased between large screen smartphones and tablets. Mental demand and effort were also significantly higher in tablet-learners compared to smartphone-learners. Considering the small difference in form-factor and design between large smartphones and small tablets this was a surprising result. There were some explanations offered in section 5.3 in the previous chapter, including the possibility of an unanticipated confounding variable identified by Koole and Ally (2009) as *psychological comfort*. This effect describes a lowering of cognitive load when an individual uses their own device rather than a device belonging to someone else. In the context of this study, it is unlikely that the results were biased by psychological comfort despite the fact that some participants were offered the use of devices that did not belong to them. When these data were removed from the analysis, there was no change in the significance of the results. However, psychological comfort is a variable that was not taken into consideration in advance of the data collection, and it may be a suitable topic for further research, perhaps indicating a qualitative approach.

[Learning Theory](#)

As stated in the limitations of the study, the learning activity used was in the cognitivist domain rather than the more currently accepted constructivist style of learning. This can be defended in the context of m-learning because when used as a distance learning tool, mobile devices are highly suited to individual study and a majority of the educational apps available for learning anatomy foster this style of learning. However, it may be of value to study mobile devices from a Human Computer Interaction (HCI) or Cognitive Load Theory (CLT) perspective when used for activities that relate to other learning styles and other modes of learning.

Sampling

The time constraints of this study necessitated the use of consecutive sampling. It might be useful to repeat the study using truly random sampling.

Qualitative Approach

Some of the results relating to the subscales of the NASA TLX, cannot be explained by numerical data alone. Further research could be considered using a qualitative methodology to explore possible reasons for high levels of reported frustration or poor performance by certain participants. Very high levels of reported physical demand in a seemingly minimally-demanding task such as interaction with an iPad might also benefit from further investigation.

6.7. REFLECTION ON INTELLECTUAL AND PROFESSIONAL DEVELOPMENT

Reflection, as defined by Schön (1983, p.68), can either refer to a process whereby a professional reflects “in action” or “on action”. Schön describes the process of reflection in action as “research in the practice context” where the reflective practitioner is required to react to events that occur in real time. Reflection “on action” considers how a practitioner may derive implications for future practice by reflecting on issues after an event has occurred. In my case, reflection in action was required throughout the entire research process, from deciding which sources to use in writing the papers during the taught component of the study, to making informed decisions about which data collection tools to use for the research component – and reacting to issues that occurred throughout the data collection process. From a personal perspective, reflections on intellectual and professional development can be categorised into two broad areas:

- General skills and knowledge that has been gained through the process of research (intellectual development).
- Recommendations for future practice that have been suggested by the results of the study (professional development).

6.7.1. Intellectual Development

Regarding intellectual development, the research process encouraged me to develop a much deeper understanding of the philosophical traditions of both education and research. Previously I had not been aware of the parallelism between the two disciplines, the intuitive rationalist approach aligning with hermeneutic enquiry and the empiricist stance aligning with an experimental post-positivist methodology. Learning about these ontological and epistemological standpoints was essential in deciding upon an appropriate theoretical and conceptual framework to my study and in choosing a suitable methodological approach. In the third year of the programme, I was concerned that the research topic and the choice of data collection tool might not be “educational” enough for a Doctorate in Education. I was, therefore, very pleased to discover that the main theoretical approaches being used were all related to learning and teaching. CLT and HCI

both being based on the work of Robert Gagné. I had not previously been aware of Gagné's input into computer science and I felt that this gave a renewed credibility to a project that had seemed to relate more strongly to technology than teaching. As a supervisor for masters projects, learning about the underpinning philosophy behind research has given me a much stronger foundation from which to advise students about the conceptual aspects of their own research projects. It also helped me to understand the nature of doctorateness, a concept that I found hard to grasp at the beginning of the process. Having completed my Master's degree in 1998, I was already aware of the difference between factual description and critical evaluation and analysis in terms of academic level. However, the background reading for this research has shown me that there are higher-level conceptual conclusions to be drawn from research and that these concepts form part of a philosophical continuum that can, in some cases, be traced back as far as human history.

From a more practical viewpoint, the choice of data collection tool proved to be a very useful experience. Because of my background in neuroscience I was keen to use physiological measures such as electroencephalography or functional Magnetic Resonance Imaging (fMRI) to assess cognitive load, but having read more about the topic I realised that there were more appropriate methods that could be applied more easily and to a larger group of participants with less ethical risk. This required me to learn about educational psychology which was a field in which I had no previous experience. In doing so, I was introduced to the work of Jean Piaget and Lev Vygotsky and other educational psychologists such as Benjamin Bloom, Robert Gagné, Jerome Bruner and John Dewey. This was very useful for me, because my route into education was via medical imaging and although I had a certain level of expertise in this field, I did not have the grounding in educational theory that would, for example, be taught in an undergraduate degree in Education Studies. My field is MRI education, but the focus was more on the scientific principles of MRI than the theoretical underpinnings of education. On reflection, I now feel that I have a more balanced skill-set in not just understanding my field, but also understanding how to facilitate learning and teaching relating to that field.

My choice of methodology proved to be a challenge. Social-sciences research often employs a qualitative approach and in the taught component of the EdD we were encouraged to consider this methodology. My research question and quantitative nature

of the data relating to cognitive load and pre/post-testing did not fit with this paradigm. As I stated in the introduction to the thesis, I had been involved in experimental research for ten years in the National Health Service – including Randomised Controlled Trials. I was therefore reasonably comfortable with the idea of using an experimental approach. However, my role in these previous projects had been largely as a research assistant, collecting data by scanning patients using MRI without any knowledge about the data analysis to follow. It was therefore really useful to have the opportunity to learn about how to set up a trial of this type, particularly looking at the ethics of research and also the data analysis, as this is something that I had never done before. Perhaps the most useful aspect of the process was discovering about the appropriate statistical tests required. As someone without a particularly strong grasp of mathematics I had reservations about using statistics and, like others in my cohort, I had considered employing a statistician to perform the necessary analysis. I quickly came to realise that this was simply not feasible. Without a grasp of the basic tenets of statistics-theory it would not be possible to construct a robust framework for the study, for example regarding the number of participants required for statistical power. From a professional perspective, my new-found knowledge of statistical tests and the assumptions necessary in deciding the correct test is useful in my professional role. For example, when supervising Master's students who intend to perform primary research, a basic knowledge of statistics is important, and I can now decide whether appropriate tests have been used with a greater degree of confidence.

The challenge of designing an on-line data collection tool was the next step in the research process, and this proved to be very time-consuming. I had already been involved in website design since the early days of the World Wide Web, and I also had experience in mobile app development, but the challenge here was to create a self-contained browser-based application that could deliver the pre and post-tests, present the learning activity, and collect data relating to the NASA TLX in one seamless package that also prevented the participants from invalidating the results of the experiment by “cheating” in the tests by using books, or failing to complete the NASA TLX scales correctly. I thought that the time invested in creating this tool, approximately six weeks' work, would be repaid by the fact that it would permit a rapid data collection process from a very large number of participants. Theoretically, all of the data could have been collected in a single

hour from hundreds of participants. This proved to be a very naïve assumption on my part. The on-line tool failed to collect data fast enough for the study. On reflection, I think this might have been due to procrastination or lack of motivation. 196 individuals agreed to take part, but without face-to-face encouragement and prompting, I think that most of them put the task on their “to do” list, but failed to follow through despite reminders. I think there is a lesson to be learnt here from a methodological perspective and perhaps even a contribution to knowledge. If the time-scale for the study had been longer, the on-line method would have eventually yielded sufficient data and would have worked well. For this reason, I have included a full description of the on-line data collection tool in Appendix A : The final benefit to me from an intellectual perspective was the opportunity to meet and share ideas with like-minded people. Discussing the studies being carried out by colleagues in the cohort, using various methodologies and investigating topics relating to both the arts and the sciences was really interesting and immeasurably useful. It taught me the importance of having a research community, and that collaboration is key when building on work that has gone before.

6.7.2. Professional Development

The examples of professional development reflected upon in the previous section were achieved in the context of the research process. In addition, there are also some implications for future practice that have been highlighted by the literature review and the results of my own research. As an author and an app developer, I have been very interested in the development of mobile devices over the last seven years and also the blurring of the boundaries between books and smartphones in delivering content. Electronic versions of books, such as Kindle editions and Apple iBook editions, are becoming increasingly popular and smartphones offer the ubiquity and portability to replace books in locations other than the home or the classroom in much the same way as they have replaced compact discs and vinyl records in delivering music.

As a web developer, I was already aware that the field of Human Computer Interaction indicates that there may be an additional cognitive load to be expected when using a computer that would not be expected when studying a book. Although no formal assumptions were made, my personal expectation was that the mobile learners would

experience a higher task load and poorer learning outcomes than the textbook learners. When I designed the app for use in the m-learning activity, I hoped that by using best-practice regarding multimedia design principles and following Mayer's (2009) Cognitive Theory of Multimedia Learning it might be possible to reduce cognitive load to a level whereby there was little or no difference between devices and books. This certainly appeared to be the case when looking at the mobile-learner group that included smartphones and tablets, but I was very surprised to discover that the smartphone-learners demonstrated a lower level of task load than the non-interactive learners. From the perspective of informing future practice, I think this result will be of benefit to me and other developers because it shows that the effort expended in creating learning materials for smartphones is empirically justified. To echo the views of Chandler (2004) and Laurillard (2008b) we are not creating the materials gratuitously, for the sake of using technology or because smartphones are a novelty, or popular with the students. We now have empirical evidence to support the use of well-designed learning apps that are optimised for smartphones.

In the words of Tabuenca, et al. (2015, p.54), a smartphone is:

“probably the only artefact co-existing with the learner in all scattered learning moments and learning contexts throughout the day”

The fact that this device may offer measurable advantages over the traditional book is a promising prospect.

7. REFERENCES

- Abdi, H.H., 2007. The Bonferonni and Šidák Corrections for Multiple Comparisons. *Encyclopedia of Measurement and Statistics*. Thousand Oaks: Sage.
- Anderson, L. W. and Krathwohl, D. R., 2000. *A Taxonomy for Learning, Teaching and Assessing: A Revision of Bloom's Taxonomy of Educational Objectives*, New York: Pearson.
- Ando, M. and Ueno, M., 2010. Analysis of the advantages of using tablet PC in e-learning, *Proceedings of the 10th IEEE International Conference on Advanced Learning Technologies*. p.2–4.
- Antonenko, P., Paas, F., Grabner, R. and Van Gog, T., 2010. Using Electroencephalography to Measure Cognitive Load. *Educational Psychology Review*, 22(4), pp.425–438.
- Apple Inc., 2013, *Market Apps with App Store Short Links* [online] Available at: <<https://itunesconnect.apple.com/WebObjects/iTunesConnect.woa>> [Accessed 25 April 2015].
- Aristotle, 2008. *De Anima*, New York: Cosimo Classics.
- Arrigo, M., Kukulska-Hulme, A., Arnedillo Sanchez, I. and Kismihok, G., 2013. Meta-analyses from a collaborative project in mobile lifelong learning, *British Educational Research Journal* 39(2), pp.222–247.
- Arsalidou, M., Pascual-Leone, J., Johnson, J., Morris, D. and Taylor, J.M., 2013. A balancing act of the brain: Activations and deactivations driven by cognitive load. *Brain and Behavior*, 3(3), pp.273–285.
- Arthur, J., Waring, M., Coe, R. and Hedges, L., 2012. *Research Methods and Methodologies in Education*, London: Sage.
- Aslanian, C. B. and Clinefelter, D. L., 2012, *Online College Students: Comprehensive Data on Demands and Preferences*, Louisville: Learninghouse/Aslanian Market Research.

Asraj, A, Freeman, M. and Chandler, P.A., 2011. Considering cognitive load theory within e- Learning environments. *Research Online – University of Wollongong*, pp.1–13.

Atkinson, R.C. and Shiffrin, R.M., 1968. Human memory: A proposed system and its control processes. In K.W. Spence and J.T. Spence, eds., *The psychology of learning and motivation*. New York: Academic Press.

Atkinson, P., 2008. A Bitter Pill to Swallow: The Rise and Fall of the Tablet Computer, *Massachusetts Institute of Technology, Design Issues*. 24(4) pp. 3-25.

Ayres, P., 2004. Research on Cognitive Load Theory and Its Design Implications for E-Learning, *Educational Technology Research and Development*, 53(3) pp. 5–13.

Ayres, P. and Sweller, J., 2005. The split-attention principle in multimedia learning. In. Mayer R. E. (Ed.) 2005, *The Cambridge handbook of multimedia learning* (pp. 134–146). New York: Cambridge University Press.

Ayres, P. and Paas, F., 2012. Cognitive load theory: New directions and challenges. *Applied Cognitive Psychology*, 26(6), pp.827–832.

Baars, B. J., 1986. *The Cognitive Revolution in Psychology*. New York: Guilford Publications Inc.

Baddeley, A.D. and Hitch, G., 1974. Working memory, *The psychology of learning and motivation: Advances in research and theory*. 8, pp. 47–89.

Baddeley, A. D., 1992. Working memory. *Science*, 255, pp.556-559.

Baddeley, A. D., 2000. The episodic buffer: A new component of working memory? *Trends in Cognitive Sciences*, 4, (11): pp.417-423.

Bloom, B., 1956. *The Taxonomy of Educational Objectives*. San Francisco: Longman.

Bohyun, K., 2013. Responsive Web Design, Discoverability, and Mobile Challenge. *Library Technology Reports*, 49(6), pp.29–40.

- Bollens, E., Rocchio, R.A., Peterson, J.E., Pollak, B., Tirpak, L. and Ward, C.M., 2014. *Understanding Responsive Web Design in Higher Education*. Working Paper: Educause Center for Analysis and Research.
- Boucheix, J. M. and Guignard, H., 2005. Which animation condition can improve text comprehension in children? *European Journal of Psychology of Education*, 20(4), pp.369-388.
- Boucheix, J.M. and Schneider, E., 2009. Static and animated presentations in learning dynamic mechanical systems. *Learning and Instruction*, 19(2), pp.112–127.
- Bowers, D., House, A. and Owens, D., 2011. *Getting Started in Health Research*. Oxford: Wiley.
- Brown, S.W., 1997. Attentional resources in timing: Interference effects in concurrent temporal and nontemporal working memory tasks. *Perception & Psychophysics*, 59(7), pp.1118–1140.
- Bruner, J. S., 1960. *The Process of Education*, Cambridge MA: Harvard University Press.
- Brünken, R., Steinbacher, S., Plass, J. L. and Leutner, D., 2002. Assessment of Cognitive Load in Multimedia Learning Using Dual-Task Methodology. *Experimental Psychology*. 49(2), pp.109–119.
- Brünken, R., Plass, J. L. and Leutner, D., 2003. Direct Measurement of Cognitive Load in Multimedia Learning. *Educational Psychologist*. 38(1), pp. 53–61.
- Brünken, R., and Seufert, T. 2009. Measuring cognitive load. In J. L. Plass, R. Moreno, and R. Brünken (Eds.), *Cognitive load theory*. New York: Cambridge University Press.
- Buchweitz, A., Mason, R.A., Tomitch, L.M.B. and Just, M.A., 2001. Brain activation for reading and listening comprehension: An fMRI study of modality effects and individual differences in language comprehension. *Life Sciences*, 2(2), pp.1–8.

Callicott, J. H., Mattay, V. S., Bertolino, a, Finn, K., Coppola, R., Frank, J. and Weinberger, D. R., 1999. Physiological characteristics of capacity constraints in working memory as revealed by functional MRI. *Cerebral Cortex*, 9(1), pp.20–6.

Card, S.K., Newell, A. and Moran, T.P., 1983. *The Psychology of Human-Computer Interaction*. Hillsdale, NJ: L. Erlbaum Assoc. Inc.

Carpenter, R. and Reddi, B., 2012. *Neurophysiology: A conceptual Approach*. London: Hodder Arnold.

Carroll, A. and Heiser, G., 2010. An analysis of power consumption in a smartphone. *Proceedings of the 2010 USENIX conference on USENIX annual technical conference*, pp.21–21.

Chan, M.Y., Haber, S., Drew, L. and Park D.C., 2014. Training Older Adults to Use Tablet Computers: Does It Enhance Cognitive Function? *The Gerontologist*, 56(03), pp.475-484.

Chandler, P., 2004. The crucial role of cognitive processes in the design of dynamic visualizations. *Learning and Instruction*, 14(3), pp.353–357.

Chapman, B., 2013. *Large Scale, Blended Learning Development: Benchmark (Research Study)*. Chapman Alliance LLC.

Chen, S., Epps, J. and Chen, F., 2011. A Comparison of Four Methods for Cognitive Load Measurement. *In Proceedings of OZCHI 2011*, pp.67–70.

Chen, C. C. and Huang, T. C., 2012. Learning in a u-Museum: developing a context-aware ubiquitous learning environment. *Computers & Education*, 59(3), pp. 873–883.

Chen, F., Zhou, J., Wang, Y., Yu, K., Arshad, S.Z., Khawaji, A. and Conway, D., 2016, *Robust Multimodal Cognitive Load Measurement (Human-Computer Interaction Series)*. New York: Springer.

Cheong, C., Bruno, V. and Cheong, F., 2012. Designing a Mobile-app-based Collaborative Learning System, *Journal of Information Technology Education*, 11. pp.97-119.

- Chu, H., 2014. Potential Negative Effects of Mobile Learning on Students' Learning Achievement and Cognitive Load — A Format Assessment Perspective. *Educational Technology and Society*, 17(1), pp.332–344.
- Clark, W. and Luckin, R., 2013. *iPads in the Classroom*. London Knowledge Lab, Institute of Education, University of London.
- Cohen, J., 1988. *Statistical Power Analysis for the Behavioral Sciences*, 2nd ed. Hillsdale: Lawrence Erlbaum.
- Commission of the European Communities, 2001, *The eLearning Action Plan*. Brussels: COM
- Compeau, D. and Higgins, C., 1995. Computer Self-Efficacy: Development of a Measure and Initial Test, *MIS Quarterly*, 19(2) pp. 189-211.
- Conroy, R.M., 2012. What hypotheses do “nonparametric” two-group tests actually test? *Stata Journal*, 12(2), pp.182–190.
- Conway, C. M. and Christiansen, M. H., 2005. Modality-constrained statistical learning of tactile, visual, and auditory sequences. *Journal of experimental psychology. Learning, memory, and cognition*, 31(1), pp.24–39.
- Cooper, G., 1998, Research into Cognitive Load Theory at the University of New South Wales. *School of Education Studies*, University of New South Wales, unpublished.
- Coulby, C., Hennessey, S., Davies, N. and Fuller R., 2011. The use of mobile technology for work-based assessment: The student experience. *British Journal of Educational Technology*, 42(2), pp.251–265.
- Cowan, N., 2001. “The Magical Number 4 in Short-Term Memory: A Reconsideration of Mental Storage Capacity,” *Behavioral and Brain Sciences*, (24:1), pp. 87-114.
- Cross, J., 2004, An Informal History of e-learning, *On The Horizon*, 12 (3), p103-110.

- Crottaz-Herbette, S., Anagnoson, R.T. and Menon, V., 2004. Modality effects in verbal working memory: Differential prefrontal and parietal responses to auditory and visual stimuli. *NeuroImage*, 21(1), pp.340–351.
- Cruzi, P. E., 2003. *A History of Modern Computing*, Cambridge MA: MIT Press.
- Dabbagh, N., 2007. The Online Learner: Characteristics and Pedagogical Implications, *Contemporary Issues in Technology and Teacher Education*, 7, pp. 217–226.
- Daniel, J., 2012. *Sampling Essentials: Practical Guidelines for Making Sampling Choices*. London: Sage.
- Dave, R., 1967. Psychomotor domain. *Proceedings of the International Conference of Educational Testing*, Berlin.
- Dearing, R. (1997) *The Report of the National Committee of Inquiry into Higher Education*. [on-line] Available at: <<http://www.leeds.ac.uk/educol/ncihe>> [accessed 13 November 2012]
- Debue, N. and van de Leemput, C., 2014. What does germane load mean? An empirical contribution to the cognitive load theory. *Frontiers in Psychology*, 5, pp.1–12.
- Deegan, R. and Rothwell, P., 2010. A Classification of M-learning Applications from a Usability Perspective. *Journal of the Research Center for Educational Technology (RCET)*, 6(1), pp.16–27.
- Deegan, R., 2015. Complex Mobile Learning that Adapts to Learners' Cognitive Load. *International Journal of Mobile and Blended Learning*, 7(1), pp.13–24.
- De Jong, T., 2009. Cognitive load theory, educational research, and instructional design: some food for thought. *Instructional Science*, 38(2), pp.105–134.
- Delazer, M., Domahs, F., Barthä, L., Brenneis, C., Lochy, A., Trieb, T., and Benke, T., 2003. Learning complex arithmetic-an fMRI study, *Cognitive brain research*, 18(1), pp.76–88.

Demouy, V. and Kukulska-Hulme, A., 2010. On the spot: using mobile devices for listening and speaking practice on a French language programme. *Open Learning: The Journal of Open and Distance Learning*, 25(3), pp.217–232.

Denzin, N.K. and Lincoln, Y.S., (2005). *Qualitative Research*. 3rd ed. London: Sage.

Descartes, R., 2011. *Meditations on the First Philosophy*. Kindle ed. London: Harper

Dickenson, D., Huxtable, R. and Parker, M. 2010. *The Cambridge Medical Ethics Workbook*. 2nd ed. Cambridge University Press.

Donnelly, R.A., 2007. *The Complete Idiot's Guide to Statistics*. 2nd ed. Oxford: Alpha.

Downes, S., 2001. Learning Objects; Resources for Distance Education Worldwide, *The International Review of Research in Open and distributed Learning*, 2(1).

Durantini, G., Gagnon, J-F., Tremblay, S. and Dehais, F., 2014. Using near infrared spectroscopy and heart rate variability to detect mental overload. *Behavioural brain research*, 259, pp.16–23.

Edmondson, W.H. and Beale, R., 2007. Projected Cognition - extending Distributed Cognition for the study of human interaction with computers. *Interacting with Computers*, 20(1), pp.128–140.

Eklund, A., Nichols, T.E. and Knutsson, H., 2016. Cluster Failure: Why fMRI inferences for spatial extent have inflated false positive rates. *Proceedings of the National Academy of Sciences of the United States of America*. 113(28), pp.7900-7905.

Ericsson Mobility Report, 2015, *On the Pulse of the Networked Society*, Stockholm: Ericsson

Faro, S. H. and Mohamed, F. B., 2010. *BOLD fMRI: A Guide to Functional Imaging for Neuroscientists*. New York: Springer.

Faul, F., Erdfelder, E., Buchner, A. and Lang, A-G., 2009. Statistical power analyses using G*Power 3.1: tests for correlation and regression analyses. *Behavior research methods*, 41(4), pp.1149–60.

- Findlater, L. and McGrenere, J., 2008. Impact of screen size on performance, awareness, and user satisfaction with adaptive graphical user interfaces. *Proceedings of the 26th SIGCHI Conference on Human Factors in Computing Systems*, pp.1247–1256.
- Fleming, M. L. and Levie, W. H., 1993. *Instructional Message Design*. 2nd ed., New Jersey: Educational technology Publications.
- Fraser, K.L., Ayres, P. and Sweller, J., 2015. Cognitive load theory for the design of medical simulations. *Simulation in Healthcare*, 10(5), pp.295–307.
- Gagné, R.M., 1985. *The Conditions of Learning and Theory of Instruction*. 4th ed.. New York: Wadsworth.
- Gagné, R. M. and Driscoll, M.P., 1988. *Essentials of Learning for Instruction*. 2nd ed. New Jersey: Prentice-Hall Inc.
- Gagné, R., Briggs, L. and Wager, W., 1998. *Principles of instructional design*. (3rd ed.). New York: Holt, Rinehart and Winston.
- Garrison, E. R., 2011. *E-Learning in the 21st Century: A Framework for Research and Practice*. 2nd ed. New York; Taylor and Francis.
- Gikas, J. and Grant, M., 2013. Internet and Higher Education Mobile computing devices in higher education: Student perspectives on learning with cellphones , smartphones & social media. *The Internet and Higher Education*, 19, pp.18–26.
- Gilbert, D., 2015. Apple's iPad Pro Finally Heralds The Post-PC Era, International Business Times, [online] Available at: < <http://www.ibtimes.com/apples-ipad-pro-finally-heralds-post-pc-era-tim-cook-2176989>> [accessed 8 April 2016].
- Ginns, P., 2005, Meta-analysis of the Modality Effect, *Learning and Instruction* 15 pp.313-331.
- Gravesetter, F.J. and Forzano, L-A, B., 2016. *Research Methods for the Behavioural Sciences*. Stamford: Cengage.
- Gray, H., 2009. *Gray's Anatomy*. London: Arcturus

Grissom, R. J. and Kim, J. J., 2012. *Effect sizes for research: Univariate and multivariate applications*. 2nd ed. New York: Taylor & Francis.

Haapalainen, E., Kim, S.J. and Dey, A.K., 2010. Psycho-physiological measures for assessing cognitive load. *Proceedings of the 12th ACM international conference on Ubiquitous computing - Ubicomp '10*, pp.301-309.

Haigh, T., 2010. "Crisis, What Crisis?" Reconsidering the Software Crisis of the 1960s and the Origins of Software Engineering. *2nd Inventing Europe/Tensions of Europe Conference, Sofia, June 17-20, 2010*, 1–36.

Hall, D.E., 2002. *The Academic Self*. Ohio: Ohio State University Press

Han, Q. and Cho, D., 2016. Characterizing the technological evolution of smartphones. *Proceedings of the 18th Annual International Conference on Electronic Commerce e-Commerce in Smart connected World - ICEC '16*, pp.1–8.

Harp, S.F. and Mayer, R.E., 1998. How seductive details do their damage: A theory of cognitive interest in science learning. *Journal of Educational Psychology*, 90(3), pp.414–434.

Hart, S.G. and Staveland, L.E., 1988. Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. In P.A. Hancock and N. Meshkati. eds., *Human mental workload*, pp.139–183.

Hart, S.G. 2006. NASA-Task Load Index (NASA-TLX); 20 Years Later, in: *50th Annual Meeting of the Human Factors and Ergonomics Society*. San Francisco: pp. 1-5.

Hart, C., 1998. *Doing a Literature Review*, London: Sage.

Hashemi, M., Azizinezhad, M., Najafi, V. and Nesari, A. J., 2011. What is m-learning? Challenges and capabilities. *Procedia - Social and Behavioral Sciences*, 30, 2477–2481.

Hawkins, D., 2014. *Biomeasurement a Student's Guide to Biological Statistics*. Kettering: Oxford University Press.

Hedges, L.V., 2012. Design of Empirical Research: In Arthur, J., Waring, M., Coe, R. and Hedges, L. eds., 2012. *Research Methods and Methodologies in Education*, London, Sage.

Hegarty, M., 2005. *Multimedia learning about physical systems*. In R. Mayer, ed. 2005. *The Cambridge Handbook of Multimedia Learning*. Cambridge: Cambridge University Press. pp. 447-465.

Heggestuen, J., 2013. One in Every 5 People in The World Own a Smartphone, One in Every 17 Own A Tablet, *Business Insider* [online] available at: <<http://www.businessinsider.com/smartphone-and-tablet-penetration-2013-10>> [Accessed 10 October 2015].

Heisenberg, W., 1998. *The Physical Principles of Quantum Theory*. New York: Dover Publications.

Herholz, S.C. and Zatorre, R.J., 2012. Musical Training as a Framework for Brain Plasticity: Behavior, Function, and Structure. *Neuron*, 76(3), pp.486–502.

Hjortskov, N., Rissen, D., Blangsted, A.K., Fallentin, N., Lundberg, U. and Søgaard, K., 2004. The effect of mental stress on heart rate variability and blood pressure during computer work. *European Journal of Applied Physiology*. 92(1-2), pp.84-9.

Höffler, T. N. and Leutner, D., 2007. Instructional animation versus static pictures: A meta-analysis. *Learning and Instruction*, (17), pp.722–738.

Hollender, N., Hofmann, C., Deneke, M. and Schmitz, B., 2010. Integrating cognitive load theory and concepts of human–computer interaction. *Computers in Human Behavior*, 26(6), pp.1278–1288.

Holley, D. and Dobson, C., 2008. Encouraging student engagement in a blended learning environment: the use of contemporary learning spaces. *Learning, Media and Technology*, 33(2), pp.139–150.

Holmes, B. and Gardner, J., 2006. *e-learning Concepts and Practice*, London, Sage.

- Huang, Y. M., Chiu, P. S., Liu, T. C. and Chen, T. S., 2011. The design and implementation of a meaningful learning-based evaluation method for ubiquitous learning. *Computers & Education*, 57(4), 2291-2302.
- Huang, Y. M., Huang, S. H. and Wu, T. T., 2014. Embedding diagnostic mechanisms in a digital game for learning mathematics. *Educational Technology Research and Development*, 62(2), 187-207.
- Human Tissue Authority, 2004. Human Tissue Act 2004. *HTA Website*, pp.1–63. [online] Available at:
<<http://www.hta.gov.uk/legislationpoliciesandcodesofpractice/legislation/humantissueact.cfm>> [accessed 2 October 2013]
- Hume, D., 2011. *The Essential Philosophical Works*. Ware: Wordsworth Editions.
- Huizenga, J. J., Admiraal, W. W., Akkerman, S. S. and Dam, G., 2009. Mobile game-based learning in secondary education: engagement, motivation and learning in a mobile city game. *Journal of Computer Assisted Learning*, 25(4), pp.332–344.
- Hurtienne, J., 2009. Cognition in HCI: An ongoing story. *Human Technology*, pp.12–28.
- Hwang, G.J., Yang, L.H. and Wang, S.Y., 2013. A concept map-embedded educational computer game for improving students' learning performance in natural science courses. *Computers & Education*, 69(March), pp.121–130.
- Ikehara, C.S. and Crosby, M.E., 2005. Assessing Cognitive Load with Physiological Sensors. *Proceedings of the 38th Hawaii International Conference on System Sciences 2005* pp.1–9.
- Johnson, L., Adams Becker, S., Estrada, V. and Freeman, A., 2015. *Horizon Report: 2015 Higher Education Edition*.
- Jones, M., Buchanan, G. and Thimbleby, H., 2003. Improving web search on small screen devices. *Interacting with Computers*, 15 (4), pp.479-495.
- Kagohara, D. M., Van der Meer, L., Ramdoss, S., O'Reilly, M. F., Lancioni, G. E., Davis, T. N. and Rispoli, M., 2012. Using iPods and iPads in teaching programs for individuals with

developmental disabilities: A systematic review. *Research in developmental disabilities*, 34(1), pp.147–156.

Kahu, E.R., 2013. Framing student engagement in higher education. *Studies in Higher Education*, 38(5), pp.758–773.

Kalyuga, S., Chandler, P. and Sweller, J., 1998. Levels of Expertise and Instructional Design. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 40(1), pp.1–17.

Kalyuga, S., Chandler, P. and Sweller, J., 1999. Managing Split-Attention and Redundancy in Multimedia Instruction, *Applied Cognitive Psychology*. 13, pp.351-371.

Kalyuga, S., Chandler, P., Tuovinen, J. and Sweller, J., 2001. When problem solving is superior to studying worked examples. *Journal of Educational Psychology*, 93, pp.579–588.

Kalyuga, S., Ayres, P., Chandler, P. and Sweller, J., 2003. The Expertise Reversal Effect. *Educational Psychologist*, 38(1), pp.23–31.

Kalyuga, S., 2011. Cognitive Load Theory: Implications for Affective Computing. In: *Proceedings 24th International Florida Artificial Intelligence Research Society Conference*. Palm Beach, Florida, United States of America, 18th – 20th May 2011. Palo Alto: Association for the Advancement of Artificial Intelligence

Kane, M. J. and Engle, R. W., 2000. Working-memory capacity, proactive interference, and divided attention: Limits on long-term memory retrieval. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 26, pp.336–358.

Kay, A. C., 1972. A personal computer for children of all ages. *Proceedings of the ACM National Conference*, 1(1), pp.1-11.

Kearney, M., Schuck, S., Burden, K. and Aubusson, P., 2012. Viewing m-learning from a pedagogical perspective, *Research in Learning Technology* 20.

- Keller, J., 2011. The slow-motion mobile campus. *The chronicle of higher education*. Available at: <<http://chronicle.com/article/The-Slow-Motion-Mobile-Campus/127380/>> [Accessed 20th December 2015].
- Kendall, J.M., 2003. Designing a research project: randomised controlled trials and their principles. *Emergency medicine journal: EMJ*, 20(2), pp.164–168.
- Khan, W., Yang, X. and Arshad, Q., 2013. Mobile Phone Sensing Systems: A Survey, *Communications Surveys and Tutorials IEEE*, 15(1) pp.402-427.
- Kim, D. and Kim, D.J., 2012. Effect of screen size on multimedia vocabulary learning. *British Journal of Educational Technology*, 43(1), pp.62–70.
- Kipper, G. and Rampolla, J., 2012. *Augmented Reality: An Emerging Technologies Guide to AR*. Waltham: Syngress (Elsevier).
- Kjeldskov, J. and Graham, C., 2003. A Review of Mobile HCI Research Methods. *Mobile HCI*, pp.317–335.
- Klopfer, E., Squire, K. and Jenkin, H., 2008. Environmental detectives – the development of an augmented reality platform for environmental simulations. *Educational Research Technology and Development*, 56(2), pp.1042–1629.
- Koole, M. and Ally, M., 2006. Framework for the Rational Analysis of Mobile Education (FRAME) model: Revising the ABCs of educational practices. *Proceedings of the International Conference on Networking, International Conference on Systems and International Conference on Mobile Communications and Learning Technologies, ICN/ICONS/MCL'06*, 2006.
- Kopcha, T. J., 2012. Teachers' perceptions of the barriers to technology integration and practices with technology under situated professional development. *Computers & Education*, 59(4), pp. 1109–1121.
- Krathwohl, D. R., 2002. A Revision of Bloom's Taxonomy: An Overview. *Theory Into Practice*, 41(4), pp. 212–218.

- Kukulska-Hulme, A. and Traxler, J., 2005. *Mobile Learning: A Handbook for Educators and Trainers*, London: Routledge.
- Kukulska-Hulme, A., Sharples, M., Milrad, M., Arnedillo Sánchez, I. and Vavoula, G., 2009. Innovation in Mobile Learning : A European Perspective. *International Journal of Mobile and Blended Learning*, 1(1), pp.13–35.
- Kukulska-Hulme, A., Sharples, M., Milrad, M., Arnedillo Sánchez, I. and Vavoula, G., 2010. Mobile learning as a catalyst for change. *Open Learning: The Journal of Open and Distance Learning*, 25(3), pp.181–185.
- Kukulska-Hulme, A., 2012. Internet and Higher Education How should the higher education workforce adapt to advancements in technology for teaching and learning? The Internet and Higher Education, 15(4), pp.247–254.
- Lau, D. and Arce, G.R., 2008, *Modern Digital Halftoning*. 2nd ed. Boca Raton: CRC Press.
- Laurillard, D., 2007. “Pedagogical forms of m-learning: Framing research questions.” In Pachler, N. ed., *Mobile learning: Towards a research agenda*. London, WLE Centre, Institute of Education, pp.153–176.
- Laurillard, D., 2008a. Technology Enhanced Learning as a Tool for Pedagogical Innovation. *Journal of Philosophy of Education*, 42(3-4), pp.521–533.
- Laurillard, D., 2008b. The teacher as action researcher: using technology to capture pedagogic form. *Studies in Higher Education*, 33(2), pp.139–154.
- Laurillard, D., Charlton, P., Craft, B., Dimakopoulos, D., Ljubojevic, D., Magoulas, G., Masterman, E., Pujadas, R., Whitley, E. A. and Whittlestone, K., 2013. A constructionist learning environment for teachers to model learning designs. *Journal of Computer Assisted Learning*, 29(1), pp.15–30.
- Land, S.M. and Zimmerman, H.T., 2015. Socio-technical dimensions of an outdoor m-learning environment: a three-phase design-based research investigation. *Educational Technology Research and Development*, 63(2), pp.229–255.

Laerd Statistics, 2016. *Laerd Statistics Premium*. [online] Available at: <<https://statistics.laerd.com>> [Accessed 2 November 2016].

Leahy, W., Chandler, P. and Sweller, J., 2003. When auditory presentations should and should not to be a component of multimedia instruction. *Applied Cognitive Psychology*, 17(4), pp.401–418.

Lee, J., Lin, L. and Robertson, T., 2012. The impact of media multitasking on learning. *Learning, Media and Technology*, 37(March 2015), pp.94–104.

Leppink, J. and van den Heuvel, A., 2015. The evolution of cognitive load theory and its application to medical education. *Perspectives on medical education*, 4(3), pp.119–27.

Lewis, T.L., Burnett, B., Tunstall, R.G. and Abrahams, P.H., 2014. Complementing anatomy education using three-dimensional anatomy mobile software applications on tablet computers, *Clinical Anatomy*, 27(3). pp.313-320.

Lim, C. P., and Khine, M. S., 2006. Managing Teachers' barriers to ICT Integration in Singapore Schools. *Journal of Technology and Teacher Education*, 14, pp.97-125.

Lin, C.L., Wang, M.J.J. and Kang, Y.Y., 2015. The evaluation of visuospatial performance between screen and paper. *Displays*, 39, pp.26–32.

Littlejohn, A., 2003. Issues in Using Online Resources, *Journal of Interactive Media in Education*. (1) pp. 1-6.

Liu, Y., Li, H. and Carlsson, C., 2010. Factors driving the adoption of m-learning : An empirical study. *Computers & Education*, 55(3), pp.1211–1219.

Liu, T.C., Lin, Y.C., Tsau, M.J. and Paas, F., 2012. Split-attention and redundancy effects on m-learning in physical environments. *Computers & Education*, 56, pp.172–181.

Lix, L.M., Keselman, J.C. and Keselman, H.J., 1996. Consequences of assumption violations revisited: A quantitative review of alternatives to the one-way analysis of variance F test. *Review of Educational Research* (66), pp.579-619.

- Longuet-Higgins, H. C., 1981. Artificial Intelligence – A New Theoretical Psychology? *Cognition*, 10(1), pp.197-200.
- Lowe, R. K., 2004. Interrogation of a dynamic visualization during learning. *Learning and Instruction*, 14, pp.257-274.
- Lupton, D., 2014. Apps as Artefacts: Towards a Critical Perspective on Mobile Health and Medical Apps, *Societies*, 4(4), pp.606-622.
- McGuire, E., 1996. Knowledge Representation and Construction in Hypermedia Environments. *Telematics and Informatics*, 13(4), pp.251-260.
- MacKenzie, I. S. and Oniszczak, A., 1998. A comparison of three selection techniques for touchpads. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (CHI '98), pp.336-343.
- Manoach, Dara S., Schlaug, G., Siewert, B., Darby, D., Bly, B., Benfield, A., Edelman, R. and Warach, S., 1997. Prefrontal cortex fMRI signal changes are correlated with working memory load, *Cognitive Neuroscience and Neuropsychology*, 8 (2), pp. 545–549.
- Manuguerra, M. and Petocz, P., 2011. Promoting student engagement by integrating new technology into tertiary education: The role of the iPad. *Asian Social Science*, 7(11), pp.61–65.
- Martin, M., 2014. *Epistemology: A Beginners Guide*, London: Oneworld Publications.
- Martin, F. and Ertzberger, J., 2013. Here and now m-learning: An experimental study on the use of mobile technology. *Computers & Education*, 68, pp.76–85.
- Mather, M., Caccioppo, J. T. and Kanwisher, N., 2013. How fMRI can inform cognitive theories. *Perspectives on Psychological Science*, 8(1), pp. 108–113.
- Mayer, R., 2009. *Multimedia Learning*. 2nd ed. Cambridge: Cambridge University Press.
- Mayer, R., 2012. *The Cambridge Handbook of Multimedia Learning*, Cambridge: Cambridge University Press.

Mayer, R.E., 2014. Incorporating motivation into multimedia learning. *Learning and Instruction*, 29, pp.171–173.

Mayfield, C.H., O'Hara, P.T. and O'Sullivan, P.S., 2012. Perceptions of a mobile technology on learning strategies in the anatomy laboratory, *Anatomy Sciences Education*, 6(2), pp.81-89.

McDonald, J.H., 2014. *Handbook of Biological Statistics*. 3rd ed. Baltimore: Sparky House Publishing.

Mehra, S., 2014. The True Cost of Building a Mobile App, *Forbes Magazine*, [online] Available at: < <http://www.forbes.com/sites/groupthink/2014/11/11/the-true-cost-of-building-a-mobile-app/2/#d3b1e677e1a5>> [Accessed 19 February 2016].

Merrill, M.D., Li, Z. and Jones, M.K., 1991. Second generation instructional design (ID2). *Educational Technology*, 30(1), pp.7–11.

Maxwell, J. 2012. *Qualitative Research Design: An Interactive Approach*. 3rd Ed. London:Sage

Miller, G. A., 1956. The Magical Number Seven, Plus or Minus Two: *The Psychological Review*, 63(1) pp.81-97.

Mittal, R., Kansal, A. and Chandra, R., 2012. Empowering developers to estimate app energy consumption. *Proceedings of the 18th annual international conference on Mobile computing and networking - Mobicom '12*, p.317.

Mohorovicic, S., 2013. Implementing responsive web design for enhanced web presence. *36th International Convention on Information & Communication Technology Electronics & Microelectronics (MIPRO)*, pp. 1206–1210.

Molina, A.I., Redondo, M. A., Lacave, C. and Ortega, M., 2014. Assessing the effectiveness of new devices for accessing learning materials: An empirical analysis based on eye tracking and learner subjective perception. *Computers in Human Behavior*, 31(1), pp.475–490.

- Monsell, S., 2003. Task Switching. *Trends in Cognitive Sciences*, 17, pp.134-140.
- Moore, D. S. and McCabe, G. P., 1999. *Introduction to the Practice of Statistics*. 3rd ed. New York: W. H. Freeman.
- Moreno, R., 2010. Cognitive load theory: more food for thought. *Instructional Science*, 38(2), pp.135–141.
- Moscovitch, M., 1992. A neuropsychological model of memory and consciousness. In L. R. Squire and N. Butters (Eds.), *Neuropsychology of memory* (2nd ed., pp. 5–22). New York: Guilford Press.
- Moscovitch, M., 1994. Cognitive resources and dual-task interference effects at retrieval in normal people: The role of the frontal lobes and medial temporal cortex. *Neuropsychology*, 8, 524–534.
- Nakano, A., 1971. Eye Movements in relation to mental activity of problem solving, *Psychologia: An International Journal of Psychology in the Orient*, 14, pp. 200-207.
- Neisser, U., 1967. *Cognitive Psychology*. Englewood Cliffs: Prentice-Hall.
- Netter, F., 2014. *Atlas of Human Anatomy*. London: Saunders.
- Nielsen, J., 1994. Enhancing the explanatory power of usability heuristics. *CHI '94: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pp.152–158.
- New Media Consortium Horizon Report, 2016. *Higher Education Edition*, Austin TX: NMC.
- NHS Employers Association, 2015. *Gender in the NHS*. London: NHS Employers Association.
- Norman, G., 2010. Likert scales, levels of measurement and the “laws” of statistics. *Advances in Health Sciences Education*, 15(5), pp.625–632.

Noyes, J., Garland, K. and Robbins, L., 2004. Paper-based versus computer-based assessment: is workload another test mode effect? *British Journal of Educational Technology*, 35(1), pp.1992–1994.

Nuyts, J., 1990. Mental processes: Studies in cognitive science. *Journal of Pragmatics*, 14(1), pp.151-154.

OFCOM, 2015. *Half of UK Homes Turn to Tablets – In Just Five Years*. [online] Available at: <<http://media.ofcom.org.uk/news/2015/five-years-of-tablets>> [Accessed 20 February 2015].

Paas, F., 1992. Training strategies for attaining transfer of problem-solving skill in statistics: A cognitive-load approach. *Journal of Educational Psychology*, 84, pp.429–434.

Paas, F.G.W.C. and van Merriënboer, J.J.G., 1993. The Efficiency of Instructional Conditions: An Approach to Combine Mental Effort and Performance Measures. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 35(4), pp.737–743.

Paas, F. G. W. C., van Merriënboer, J. J. G. and Adam, J. J., 1994. Measurement of cognitive load in instructional research. *Perceptual and Motor Skills*, 79, 419–430.

Paas, F., Tuovinen, J. E., Habbers, H. and Van Gerven, P. W. M., 2003. Cognitive Load Measurement as a Means to Advance Cognitive Load Theory. *Educational Psychologist*, 38(1), pp.63–71.

Paas, F. and Sweller, J., 2011. An Evolutionary Upgrade of Cognitive Load Theory: Using the Human Motor System and Collaboration to Support the Learning of Complex Cognitive Tasks. *Educational Psychology Review*, 24(1), pp.27–45.

Paas, F. and Ayres, P., 2014. Cognitive Load Theory: A Broader View on the Role of Memory in Learning and Education. *Educational Psychology Review*, pp.1–5.

Pachler, N., Bachmair, B. and Cook, J., 2010. *Mobile Learning, Structures, Agencies, Practices*, New York: Springer.

Park, B., Moreno, R., Seufert, T. and Brünken, R., 2010. Does cognitive load moderate the seductive details effect? A multimedia study. *Computers in Human Behaviour*, 27(2011), pp.5-10.

Paus, T., 2005. Mapping brain maturation and cognitive development during adolescence. *Trends in Cognitive Sciences*, 9(2), pp.60–68.

Paivio, A., 1990. *Mental Representations: A Dual Coding Approach*, Oxford: Oxford University Press.

Pavlov, I.P., 1927, *Conditioned Reflexes*. London: Oxford University Press

Penney, C. G. 1989. Modality effects and the structure of short-term memory. *Memory and Cognition*, 17, pp.398-442.

Pereira, A., Miller, T., Huang, Y-M., Odell, D. and Rempel, D., 2013. Holding a tablet computer with one hand: effect of tablet design features on biomechanics and subjective usability among users with small hands. *Ergonomics*, 56(9), pp.1363–75.

Perry, M., O'Hara, K., Sellen, A., Harper, R. and Brown, B., 2001. Dealing with mobility: understanding access anytime, anywhere. *ACM Transaction on Computer-Human Interaction*, 8(4), pp. 323–347.

Pett, D. and Wilson, T., 1996. Color research and its application to the design of instructional materials. *Educational Technology Research and Development*, 44, 3, pp. 19–35.

Piaget, J., 1969. *The Psychology of the Child*. New York: Perseus.

Piaget, J., 2001. *La Psychologie de l'Intelligence (The Psychology of Intelligence)*, Abingdon: Routledge.

Plato, *Meno*, transl. Guthrie, W. K. C., 1973. *Plato: Collected Dialogues*, Princeton: Princeton University Press.

- Pochon, J.B., Levy, R., Poline, J. B., Crozier, S., Lehericy, S., Pillon, B., Deweer, B., Le Bihan, D. and Dubois, B., 2001. The Role of Dorsolateral Prefrontal Cortex in the Preparation of Forthcoming Actions: an fMRI Study. *Cerebral Cortex*, 11(3), pp.260–266.
- Price, S. and Rogers, Y., 2004. Let's get physical: the learning benefits of interacting in digitally augmented physical spaces. *Computers & Education*, 43, pp.137-151.
- Ramirez, S., Liu, X., Lin, P-A., Suh, J., Pignatelli, M., Redondo, R.L., Ryan, T.J. and Tonegara, S., 2013. Creating a false memory in the hippocampus. *Science*, 341(6144), pp.387–391.
- Raptis, D., Tselios, N., Kjeldskov, J. and Skov, M. B., 2013. Does Size Matter? Investigating the Impact of Mobile Phone Screen Size on Users' Perceived Usability, Effectiveness and Efficiency. *Proceedings of the International Conference on Human-Computer Interaction with Mobile Devices and Services (Mobile HCI '13)*, pp.127–136.
- Razali, N.M. and Wah, Y.B., 2011. Power comparisons of Shapiro-Wilk , Kolmogorov-Smirnov, Lilliefors and Anderson-Darling tests. *Journal of Statistical Modeling and Analytics*, 2(1), pp.21–33.
- Reeves, B. and Nass, C., 1996. *The media equation: How people treat computers, television, and new media like real people and places*. New York: Cambridge University Press.
- Rodriguez-Moreno, D. and Hirsch, J., 2006. The dynamics of deductive reasoning: An fMRI investigation. *Neuropsychologia*, 47(4), pp.949–961.
- Rossing, J.P., Miller, W.M., Cecil, A.K. and Stamper, S.E., 2012. iLearning : The future of higher education ? Student perceptions on learning with mobile tablets. *Journal of the Scholarship of Teaching and Learning*, 12(2), pp.1–26.
- Rubio, S., Díaz, E., Martín, J. and Puente, J., 2004. Evaluation of Subjective Mental Workload: A Comparison of SWAT, NASA-TLX, and Workload Profile Methods., 53(1), pp.61–86.
- Rugg, G., and Petre, M., 2007. *A Gentle Guide to Research Methods*. 1st ed. Maidenhead: Open University Press.

- Salkind, N.J., 2008. *Statistics for People Who Think They Hate Statistics*, London: Sage.
- Sarrab, M., Elbasir, M. and Alnaeli, S., 2016. Towards a quality model of technical aspects for m-learning services: An empirical investigation. *Computers in Human Behavior*, 55, pp. 100–112.
- Savoji, A.P., Hassanabadi, H. and Fasihipour, Z., 2011. The modality effect in learner-paced multimedia learning. *Procedia - Social and Behavioral Sciences*, 30, pp.1488–1493.
- Schmidt-Weigand, F. and Scheiter, K., 2011. The role of spatial descriptions in learning from multimedia. *Computers in Human Behavior*, 27(1), pp.22–28.
- Schnotz, W. and Kürschner C., 2007. A Reconsideration of Cognitive Load Theory. *Educational Psychology Review*, 19, pp.469-508.
- Schomer, D. L. and Lopes da Silva, F., 2011. *Niedermeyer's Electroencephalography: Basic Principles, Clinical Applications, and Related Fields*. London: Lipincott, Williams and Wilkins.
- Schön, D.A., 1983. *The Reflective Practitioner: How Professionals Think in Action*. New York: Basic Books.
- Schwan, S. and Riempp, R. 2004. The Cognitive Benefits of Interactive Videos: Learning to Tie Nautical Knots, *Learning and Instruction*, 14, pp.293-305.
- Seong, D.S.K., 2006. Usability guidelines for designing mobile learning portals. *Proceedings of the 3rd international conference on Mobile technology, applications & systems - Mobility '06*, p.25.
- Setz, C., Arnrich, B., Schumm, J., La Marca, R. and Tröster, G., 2010. Discriminating Stress from Cognitive Load Using a Wearable EDA Device. *Transactions on Information Technology in Biomedicine*, 14(2), pp.410–417.

- Shadieff, R., Hwang, W-Y., Huang, Y-M. and Liu, T-Y., 2015. The Impact of Supported and Annotated Mobile Learning on Achievement and Cognitive Load. *Educational Technology & Society*, 18(4), pp.53–69.
- Shapiro, S. S. and Wilk, M.B., 1965. An Analysis of Variance Test for Normality (Complete Samples) *Biometrika*, 52 (3/4). pp. 591-611.
- Sharples, M., Taylor, J. and Vavoula, G., 2007. A Theory of Learning for the Mobile Age A Theory of Learning for the Mobile Age (pre-print). *The Sage Handbook of Elearning Research*, Thousand Oaks CA: Sage.
- Shin, D. H., Shin, Y. J., Choo, H. and Beom, K., 2011. Smartphones as smart pedagogical tools: Implications for smartphones as u-learning devices. *Computers in Human Behavior*, 27(6), pp.2207–2214.
- Shirer, M. 2015. *As Tablets Slow and PCs Face Ongoing Challenges, Smartphones Grab an Ever-Larger Share of the Smart Connected Device Market Through 2019, According to IDC*, [online] Available at: <<http://www.idc.com/getdoc.jsp?containerId=prUS25500515>> [Accessed 23rd March 2016].
- Shellock, F. G., 2007. *Reference Manual for Magnetic Resonance Safety, Implants and Devices*. Los Angeles: Biomedical Research Publishing Group.
- Shuster, D.P. and Powers, W.J., 2005, *Translational and Experimental Clinical Research*. Philadelphia: Lippincott, Williams and Wilkins.
- Singh, S., 2005, *Big Bang*, London: Harper Collins.
- Skinner, B.F., 1988. *About Behaviourism*. New York: Random House.
- Smørðal, O. and Gregor, J., 2005. Knowmobile. In: A. Kukulska-Hulme and J. Traxler, 2005. *Mobile Teaching and Learning. Mobile Learning: A Handbook for Educators and Trainers*, London: Routledge.
- Snell, J., 2013. Flexible everything: Getting responsive with web design. *Computers in Libraries*, 33(3), pp.12–16.

- Sølvberg, A. M. and Rismark, M., 2012. Learning spaces in m-learning environments. *Active Learning in Higher Education*, 13(1), pp. 23–33.
- Souleles, N., Savva, S., Watters, H., Annesley, A. and Bull, B., 2014. A phenomenographic investigation on the use of iPads among undergraduate art and design students. *British Journal of Educational Technology*, (2014), pp.1-11.
- Spencer, L., Jakobsen, M., Shah, S. and Cairns, G., 2013. Minimum required angular resolution of smartphone displays for the human visual system. *Journal of the Society for Information Display*, 21(8), pp.352–360.
- Subramanian, C., 2012. New Study Finds iPads in the Classroom Boost Test Scores. *Time Tech*. 22 Feb 2012. Available at: <<http://techland.time.com/2012/02/22/new-study-finds-ipads-in-the-classroom-boost-test-scores>> [accessed 23 April 2013].
- Sun, P.C., Tsai, R. J., Finger, G., Chen, Y-Y. and Yeh, D., 2008. What drives a successful e-Learning? An empirical investigation of the critical factors influencing learner satisfaction. *Computers & Education*, 50(4), pp.1183–1202.
- Sung, E. and Mayer, R.E., 2013. Multimedia learning with mobile devices and desktop computers: An experimental test of Clark 's methods-not-media hypothesis. *Computers in Human Behavior*, 29(3), pp.639–647.
- Sung, Y., Chang, K. and Yang, J., 2015. How effective are mobile devices for language learning? *Educational Research Review*, 16, pp.68–84.
- Sung, Y., Chang, K. and Liu, T., 2016. The effects of integrating mobile devices with teaching and learning on students' learning performance: A meta-analysis and research synthesis. *Computers & Education*, 94, pp.252–275.
- Sweller, J., 1989. Cognitive technology: Some procedures for facilitating learning and problem solving in mathematics and science. *Journal of Educational Psychology*, 81(4), pp.457–466.

- Sweller, J., 1994. Cognitive Load Theory, *Learning Difficulty and Instructional Design, Learning and Instruction*. 4, pp.295–312.
- Sweller, J., van Merriënboer, J.J.G. and Paas, F.G.W.C., 1998. Cognitive Architecture and Instructional Design. *Educational Psychology Review*, 10(3), pp.251–296.
- Sweller, J., Ayres, P. and Kalyuga, S., 2011. *Cognitive Load Theory*, New York: Springer.
- Sweller, J., 2018. *Discussion about criticisms of CLT*. [email] (Personal communication, 14 February 2018).
- Tabuenca, B., Kalz, M., Drachler, H. and Specht, M., 2015. Time will tell: The role of m-learning analytics in self-regulated learning. *Computers & Education*, 89, pp.53–74.
- Tarmizi, R. and Sweller, J., 1988. Guidance during mathematical problem solving. *Journal of Educational Psychology*, 80, pp.424-436.
- TEL Report, 2012. *System Upgrade – Realising the Vision for UK education, 2012. A report from the ESRC/EPSRC Technology Enhanced Learning Research Programme*. London Knowledge Lab.
- Terras, M. and Ramsay, J., 2012. The five central psychological challenges facing effective m-learning. *British Journal of Educational Technology*, 43(5), pp. 820–832.
- The Society of Radiographers, 2009. *Effect of Agenda for Change on Career Progression of the Radiographic Workforce*. London: Society of Radiographers.
- Traxler, J., 2007. Defining, discussing and evaluating m-learning: The moving finger writes and having writ. *The International Review of Research in Open and Distance Learning*, 8(2). pp.18-26.
- Traxler, J., 2009. Learning in a Mobile Age. *International Journal of Mobile and Blended Learning*, 1(1), pp.1-12.

- Trinder, J., 2005. Mobile Technologies and Systems. In: Kukulska-Hulme, A. and Traxler, J., 2005. *Mobile Teaching and Learning. Mobile Learning: A Handbook for Educators and Trainers*, London: Routledge.
- Trowler, V. and Trowler, P., 2010. *Student engagement evidence summary*. York: Higher Education Academy.
- Tsao, Y-C., Wittlieb, E., Miller, B. and Wang, T-G., 1983. Time estimation of a secondary event. *Perceptual and Motor Skills*, 57(3), pp.1051–1055.
- Tversky, B., Bauer-Morrison, J. and Bétrancourt, M., 2002. Animation: Can it facilitate? *International Journal of Human–Computer Studies*, (57), pp.247– 262.
- Twining, P. and Evans, D., 2005. Should there be a future for Tablet PCs in schools? *Journal of Interactive Media in Education*. 2005(20), pp.1-18.
- Van Gog, T., and Paas, F. 2008. Instructional efficiency : Revisiting the original construct in educational research. *Educational Psychologist*, 43, pp.16–26.
- van Merriënboer, J. J. G. and Ayres, P., 2004. Research on Cognitive Load Theory and Its Design Implications for E-Learning, *Educational Technology Research and Development*, 53(3), pp. 5-13.
- van Merriënboer, J. J. G. and Sweller, J., 2010. Cognitive load theory in health professional education: design principles and strategies. *Medical Education*, 44(1), pp.85–93.
- Vavoula, G., Pachler, N. and Kukulska Hulme, A., 2009, *Researching Mobile Learning: Frameworks, Tools, and Research Designs*, London: Verlag.
- Vogel-Walcutt, J.J., Gebrim, J.B., Bowers, C., Carper, T.M. and Nicholson, D., 2011. Cognitive load theory vs. constructivist approaches: Which best leads to efficient, deep learning? *Journal of Computer Assisted Learning*, 27(2), pp.133–145.
- Vygotsky, L., 1986. *Thought and Language*, Massachusetts: MIT Press.

- Wang, M. and Shen, R., 2012. Message design for m-learning: Learning theories, human cognition and design principles. *British Journal of Educational Technology*, 43(4), pp.561–575.
- Waring, R., 2012. Finding Your Theoretical Position: In J. Arthur, M. Waring, R. Coe and L. Hedges, eds. 2012. *Research Methods and Methodologies in Education*, London: Sage
- Watson, N., 2013, *Getting to Know iTunes U Webcast* [online] Available at: <<http://euro.edu.apple.kuluvalley.com/en/event3>>, [accessed 18 April 2013].
- Watson, J.B., 1931. *Behaviourism*. New York: Kegan Paul, Trench, Trubner & Co.
- Waugh, A. and Grant, A. 2014. Ross and Wilson Anatomy and Physiology in Health and Illness. London: Churchill Livingstone.
- Whelan, R. R., 2007. Neuroimaging of cognitive load in instructional multimedia. *Educational Research Review*, 2(1), pp. 1–12.
- White, S., 2000. In H. Fry, S. Ketteridge and S. Marshall. eds. *A Handbook for Teaching & Learning in Higher Education*. London: Kogan.
- Westbrook, C.A., Talbot, J. S. and Kaut-Roth, C., 2011. *MRI in Practice*, 4th ed. Oxford: Wiley-Blackwell.
- Wong, A., Leahy, W., Marcus, N. and Sweller, J., 2012. Cognitive Load Theory, the transient information effect and e-learning. *Learning and Instruction*, 22, pp.449-457.
- Younus, A., Qureshi, M. A., Saeed, M., Ali, S. A. and Touheed, N., 2011. IBook : Interactive and Semantic Multimedia Content Generation for eLearning, *International Journal of Computer Science* 8(3), pp. 438–443.
- Young, J. R., 2010. Will the iPad Be Able to Bring the Internet to the Beanbag Chair? *The Chronicle of Higher Education*, 56 (31).

Young, J.Q., van Merriënboer, J.G., Durning, S. and Cate, O. T., 2014. Cognitive Load Theory: implications for medical education: AMEE Guide No. 86. *Medical Teacher*, 36(5), pp.371–384.

Zhang, P. and Galletta, D., 2015. *Human Computer Interaction and Management Information Systems*. Abingdon: Routledge.

Zhao, L., 2011. The Social Dimension of Distance Learning by Interactive Television: A Qualitative Study, *Current Issues in Education*, 14(2).

Zydney, J.M. and Warner, Z., 2015. Mobile apps for science learning: Review of research. *Computers & Education*, 94, pp.1–17.

8. APPENDICES

Appendix A : The Online Data Collection Method (Discontinued)

The following section describes and seeks to justify the use of an online data collection tool. The method was discontinued after a number of months as it became apparent that there were some drawbacks, particularly in relation to the time taken to gain sufficient data for the study. This section is therefore provided in the interests of contribution to methodological knowledge. The revised data collection method is explained in the section beginning on page 135.

Sampling technique for the originally intended online data collection method

For the originally proposed online data-collection method, convenience sampling was considered. According to Gravesetter and Forzano (2016), this type of sampling is probably the most commonly used sampling method in the behavioural sciences such as psychology. However, convenience sampling is not considered ideal in studies where generalisation to a target population or statistical significance is intended. Convenience sampling was considered for this study as it would permit recruitment of the requisite number of participants in the time available for data collection and because there was no pre-existing census of potential participants from which to draw a random sample. Due to the online nature of the data collection it was thought that main flaw of convenience sampling could be minimised, namely that this approach would by nature attract a wide and varied sample of the target population, geographically and by age and gender. Participants were invited from the researcher's professional networks (Facebook MRI in Practice group and Linked In) and students from the researcher's own institution, but not under his direct tutelage. Ethical approval was sought and granted to invite students from another university (undertaking medical-imaging courses).

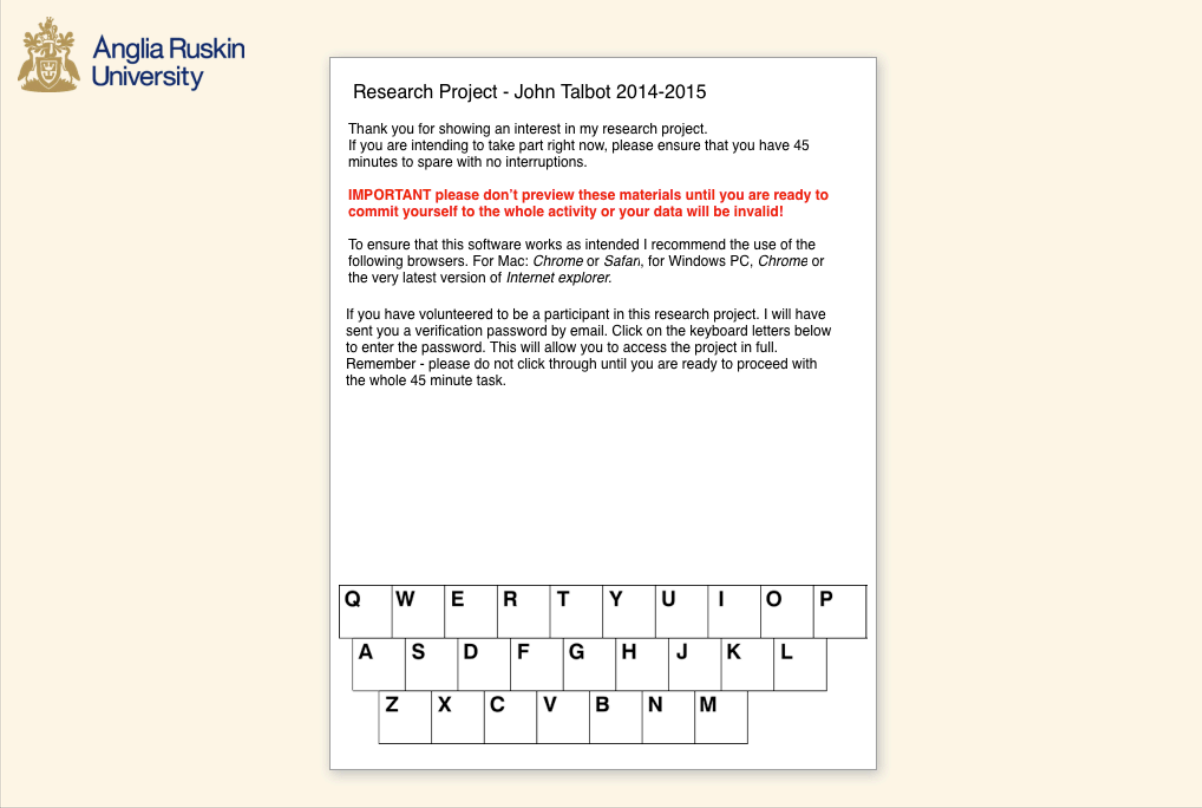
Invitations were sent out to approximately 2,400 individuals, 196 of whom agreed to take part in the online data collection. However, after three months of data collection, only 14 participants had attempted the online tasks and some of the tasks were not completed fully or were completed incorrectly. This led to concerns over bias, and also completion time for the study as well as the inherent non-generalisability of results gathered from small convenience samples. It was thought that there was a risk of attrition-bias and self-selection bias, namely that individuals with a particular interest in technologically-mediated learning


might have been more motivated to participate. Compeau and Higgins (1995) first described a phenomenon known as computer self-efficacy (CSE) defined as “an individual judgement of one’s capability to use a computer” (Compeau and Higgins, 1995, pp.129). The concern here is that participants having a low CSE were discouraged from completing the online data collection activity, resulting in data being only collected from participants having a high CSE. This scenario would have introduced a potential cause of bias, as the target population cannot be assumed to have a high level of CSE.

The second concern was completion time for data collection. By logging the number of participants undertaking the online data-collection over the first three months of the trial (and extrapolating the declining rate of participation) it became clear that the time required to complete the data collection would have been in excess of three years. This would have made the study impractical due to the time constraints of the doctoral academic process. Reasons for the poor response rate were not ascertained but reminder emails were sent out, and the response rate did not improve. After a consultation with the supervisory team, it was thought that the participants might have thought that they were not sufficiently anonymous, having been assigned a participant ID number to provide during the online tasks. This is a feasible explanation because participants are likely to have felt uncomfortable about their lack of knowledge in human anatomy, an area that was (in many instances) related to their profession.

The application created for data collection was written by the researcher and hosted on his web server. The application featured twelve screens that could be navigated consecutively in a web browser. These sections contained the following features:

Screen 01 – Welcome and Log-in:



 Anglia Ruskin University

Research Project - John Talbot 2014-2015

Thank you for showing an interest in my research project.
If you are intending to take part right now, please ensure that you have 45 minutes to spare with no interruptions.

IMPORTANT please don't preview these materials until you are ready to commit yourself to the whole activity or your data will be invalid!

To ensure that this software works as intended I recommend the use of the following browsers. For Mac: *Chrome* or *Safari*, for Windows PC, *Chrome* or the very latest version of *Internet explorer*.

If you have volunteered to be a participant in this research project. I will have sent you a verification password by email. Click on the keyboard letters below to enter the password. This will allow you to access the project in full.
Remember - please do not click through until you are ready to proceed with the whole 45 minute task.

Q	W	E	R	T	Y	U	I	O	P
A	S	D	F	G	H	J	K	L	
Z	X	C	V	B	N	M			

Screen shot of screen 01 of the online data collection tool

This opening screen thanked the participant for agreeing to take part in the trial and explained that they should not begin the activity unless they had enough free time to undertake the entire activity. There was a virtual keyboard presented on screen to allow the participant to enter a password provided to them by email. A user password was required to prevent unauthorised data entry by non-participants. The reason that the entire activity was required to be undertaken in one session was to prevent participants pausing to search for answers to the pre and post tests and then returning to the activity. This would have invalidated the results as the pre/post test scores would not have been undertaken from memory.

Screen 02 – participant information sheet:

Anglia Ruskin University

Participant Information Sheet (scroll down to read)

My name is John Talbot I am a senior lecturer in magnetic resonance imaging, research and applied anatomy at Anglia Ruskin University in Cambridge. I am currently undertaking a research project entitled

"An Empirical Study to Assess the Impact of Mobile Touch-Screen Learning on User Information Load".

Thank you for showing interest in taking part in my research project. Here are the answers to some questions you may have about participating.

Please read this document carefully and ensure that you have read the consent section at the end.

Introduction

There is a new branch of learning, known as mobile learning, that uses mobile phones and touch-screen computers as learning tools. It allows the student to study wherever they happen to be, and permits the use of images, sound and user interaction.

Previous research suggests that teachers sometimes design their teaching materials to suit the features offered by this new technology rather than to address the way students actually learn. This is of concern because meaningful learning may be hampered rather than enhanced by the use of the technology.

A research study into how these devices affect learning (for better or worse) would be therefore useful when designing learning activities.

What is the purpose of the research?

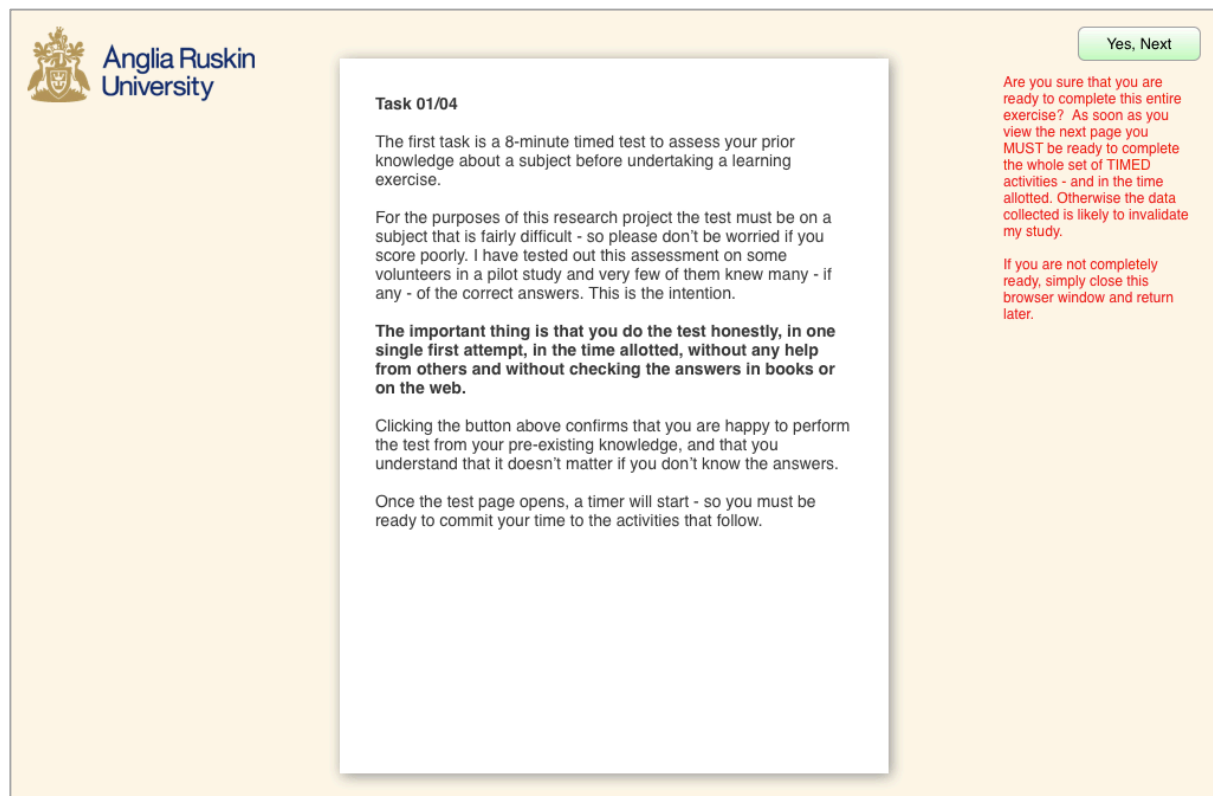
Previous studies have shown that certain types of learning activity require

Confirm

Screenshot of screen 02 of the online data collection tool.

This screen featured a scrollable participant information sheet (see appendix XX). On clicking the confirm button a message appears asking the participant to confirm that they have read the entire information sheet and to confirm consent to take part in the study. On clicking the consent button the application opened screen 03.

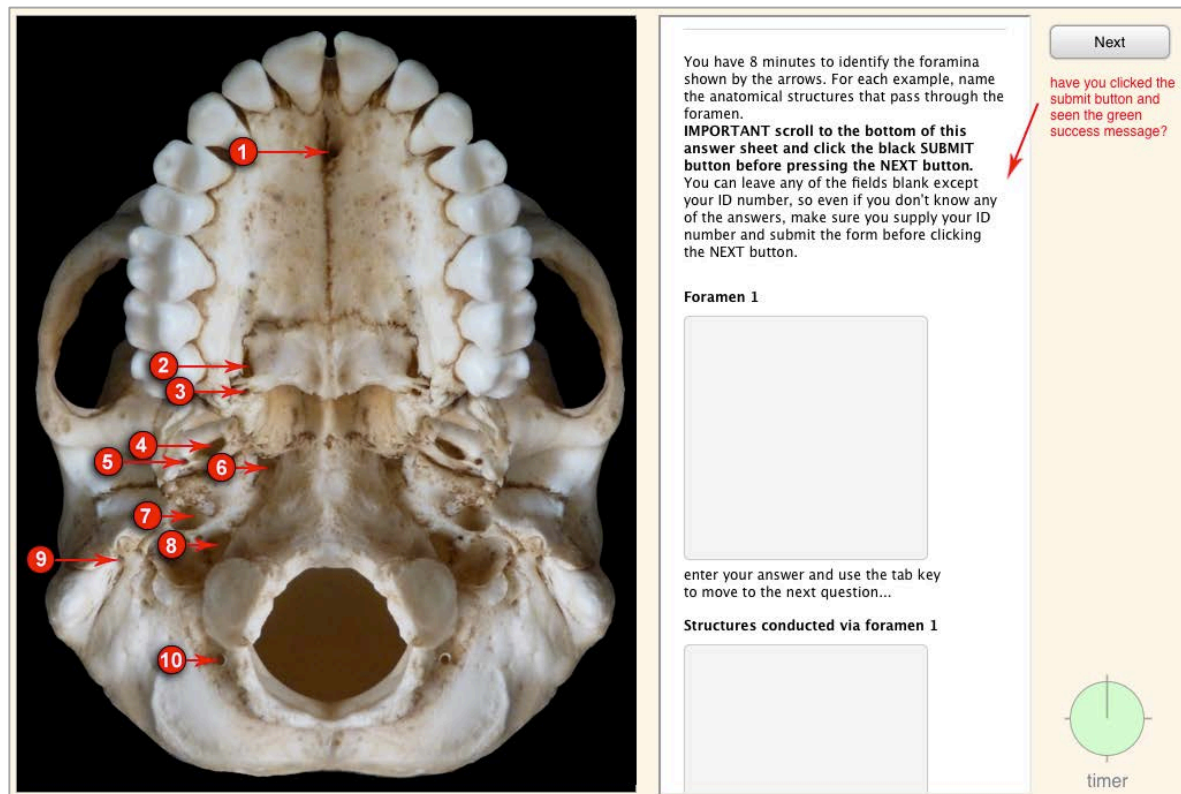
Screen 03 – Pre-test explanation screen:



Screenshot of screen 03 of the online data collection tool.

Screen 03 explained that the participant was about to be presented with a pre-test and that the test was intentionally difficult (i.e. that it was anticipated that they would know the answers). This was articulated to avoid the possibility of them thinking that they should know the answers and be discouraged from undertaking the task. The screen instructions explained that it was important that they take the test from their own knowledge rather than seeking help from others or looking the answers up in books or websites. On clicking the start-button a confirmation message appeared (see fig. 3.4) asking the participant to confirm that they were ready to take the entire data-collection task and giving them the opportunity to return later if they were not ready. This is the last point that it was possible to return to the data collection without the possibility of invalidating the results, because the following screen shows the anatomy to be learnt.

Screen 04 – Pre-test on the Anatomy of the Skull Base:



Screenshot of screen 04 of the online data collection tool.

Screen 04 featured the pre-test. The pre-test consisted of a photograph showing a replica human skull-base (resin cast and indistinguishable from the genuine article) with ten numbered arrows indicating the foramina of the skull base. These are windows in the bone that allow the passage of nerves and blood vessels. To the right of the diagram was a web-form containing 10 fields that required the participants to enter the names of the foramina and another 10 fields that required the participants to identify the structures passing through each foramina. Each field was a text-box that allowed the participants to write a number of answers within each if required. The highest possible score for participants correctly identifying all structures was 36 (10 foramina and 26 nerves and blood vessels). There was also a field to allow the participant to provide their ID number to identify them and a button to submit their answers at the end of the test. The form was configured to allow fields to be left blank if the participant did not know the answer.

There was a built in timer of 8 minutes (see fig. 3.5) to ensure that all participants had equal amounts of time to complete the task. At the end of the 8 minute timed exercise the image

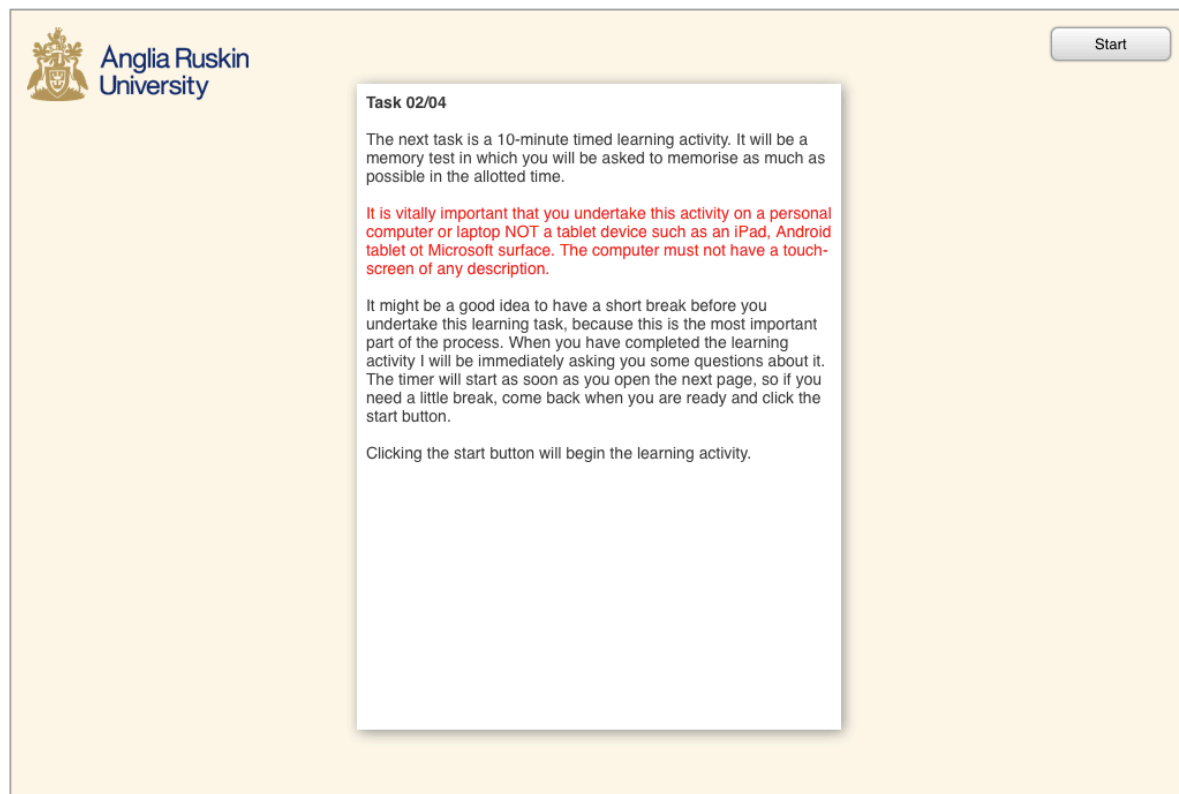
was blanked out to prevent further answering, but the form was left visible to allow the participants to submit their answers. As this intended data-collection method was remote and therefore unsupervised by the researcher, there were a few web-design considerations required to ensure that data were collected without any technical issues and without presenting the participants with an opportunity to invalidate the pre-test data, for example had the participant opted to leave the form open and find the answers from another source. This would have given misleading data as it would not have represented the participants' knowledge before any learning activity had occurred. This scenario was deemed unlikely because without the photograph as a reference the participant would have found it difficult to remember which numbered form-field related to which specific foramen of the skull. The timing of 8 minutes was chosen as this allowed enough time to write the answers, but not enough time to search for the answers online or in a textbook.

Web-browsers typically feature a back-button but for this task, one-way navigation was required to prevent participants revisiting the pre-test having undertaken the following learning activity. This would also have invalidated the pre-test data for the same reason mentioned above. To maintain validity, navigation through the task was only made possible by using the controls in the application itself, and was only configured for one-way forward navigation through the stages of the task.

In some web-forms, the text fields are designed so that pressing the return key submits the form. To avoid the potential pitfall of a participant unwittingly submitting the form before having entered all of the answers, the text boxes were formatted to prevent use of the return key submission. To prevent the participant from moving onto the next section of the task without having submitted the form, an alert message was employed to remind the participant that the submit button should be pressed before progressing to the next stage. On clicking the submit button, the test-answers were automatically stored in an online database, and also sent to a purpose-designed email address as a back-up in case of database failure.

There were no indications given on the pre-test page that there was to also be a post-test as it was thought that this would have allowed participants to gain an unfair advantage during the subsequent learning activity (for example, writing down the diagram labels rather than trying to learn them by rote).

Screen 05 – Learning activity explanation screen:



Screenshot of screen 05 of the online data collection tool.

Screen 05 explained that the next step of the process was a learning activity.

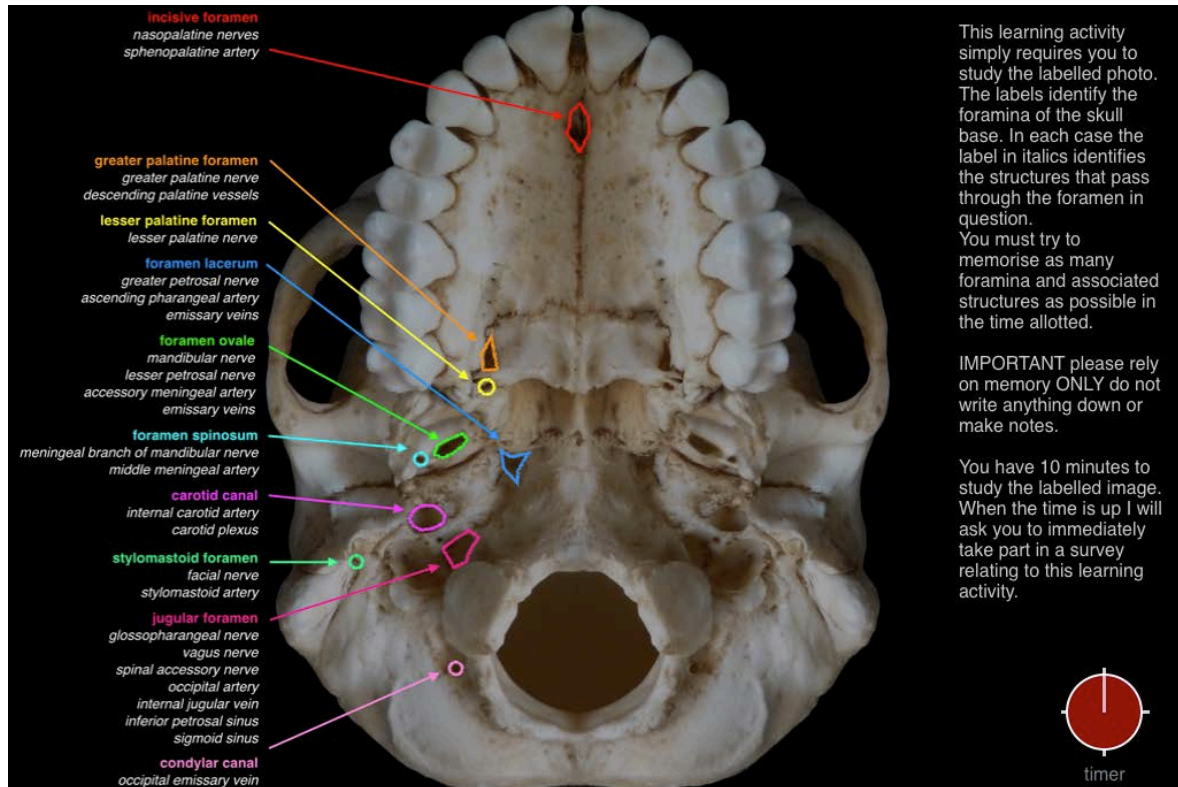
The control group version of the app explained that the activity must be taken on the computer rather than a touch screen device. The web app was designed to interrogate the users screen size and device to prevent participants from accessing the touch-screen materials on a PC and vice versa. In the event that a participant attempted to access the m-learning materials on a PC a blank screen was displayed with a message stating that the materials must be accessed on a smartphone or tablet.

The experimental-group version of the app (accessed via a different link) explained that the activity must be taken on a touch-screen device rather than on the PC and a link provided to access this.

This originally proposed online data collection method intended to replicate a non-digital learning material by presenting a photograph on the computer screen, ecological validity

was improved in the reconfigured data collection method where participants were provided with a labelled photograph rather than having to use a computer.

Screen 06 : Learning activity (control group version):



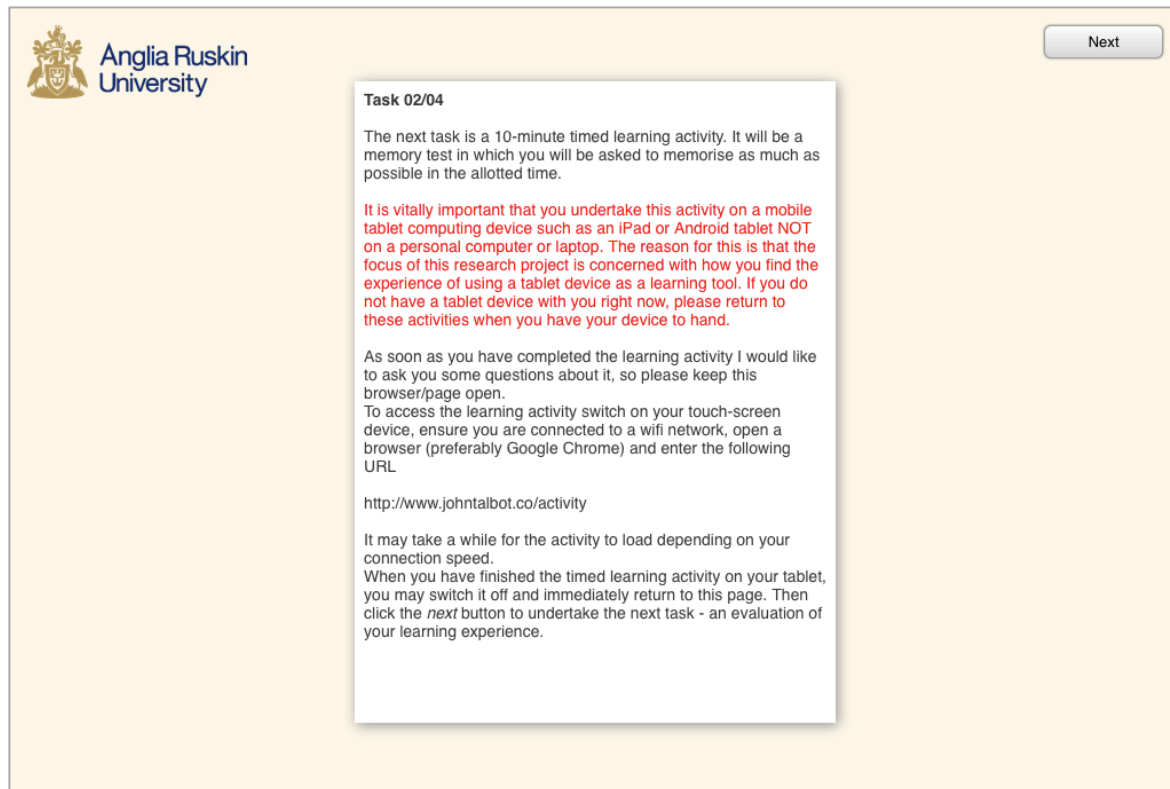
Screenshot of screen 06 of the online data collection tool (the learning activity)

This section of the data collection tool differed according to whether participants were in the control group or the experimental group.

For the control group this screen featured the non-interactive learning activity and was intended to replicate a labelled textbook photograph showing the base of skull. The captions were placed as close to the associated anatomical features to reduce the split attention effect (Nielsen, 1994; Mayer, 2009). Participants were given ten minutes to learn (memorise) as many of the structures as possible. After ten minutes had elapsed on the timer, the screen automatically closed. This ensured that all participants had the same experience and controlled for a time variable in knowledge acquisition. If participants had used the browser back button in order to gain more time to learn, the app was configured not to permit this.

For the experimental group a different screen was displayed in the browser shown on the following page.

Screen 06 - Learning activity link (experimental group version):

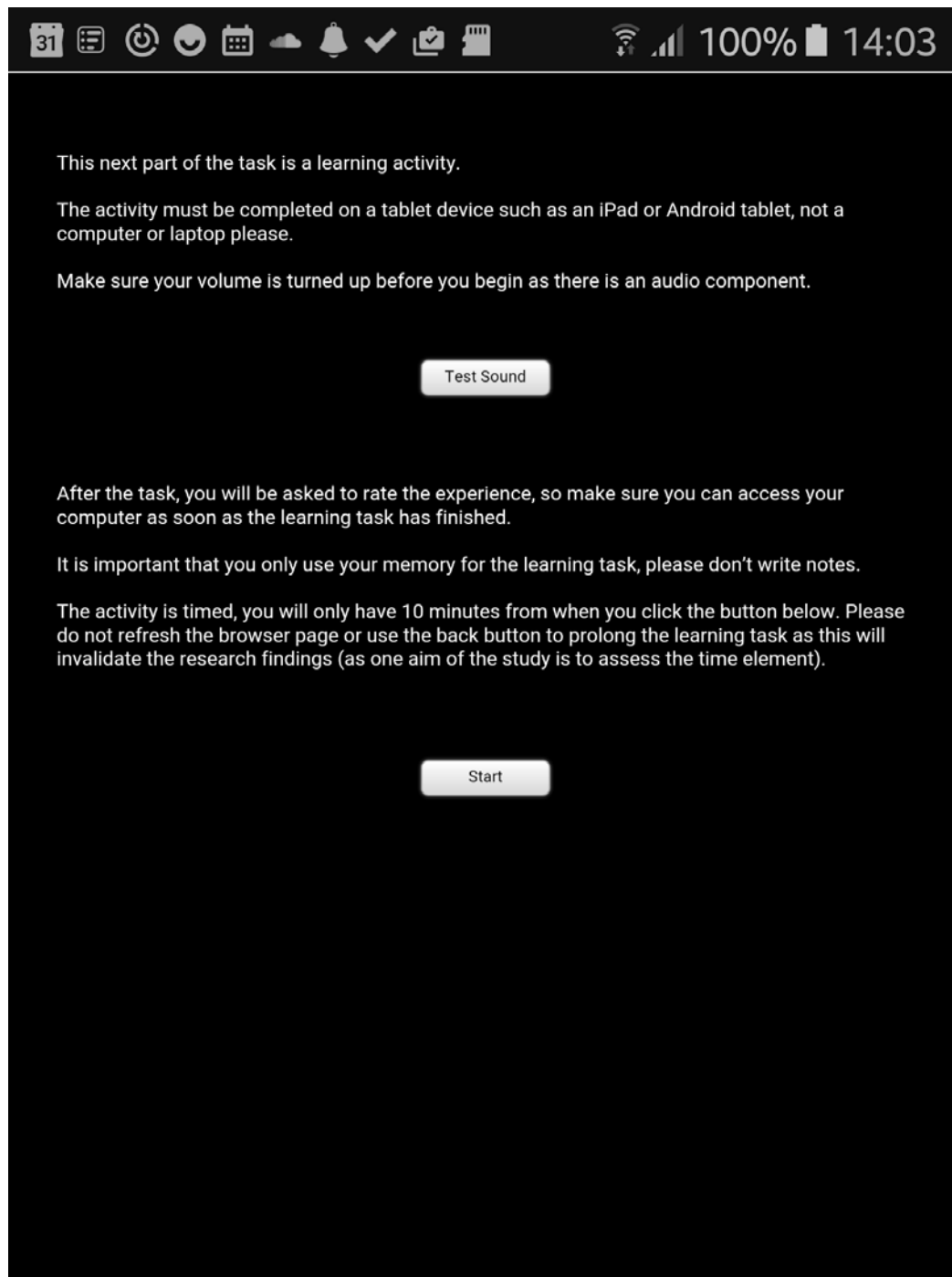


Screenshot of screen 06 of the online data collection tool for the experimental group.

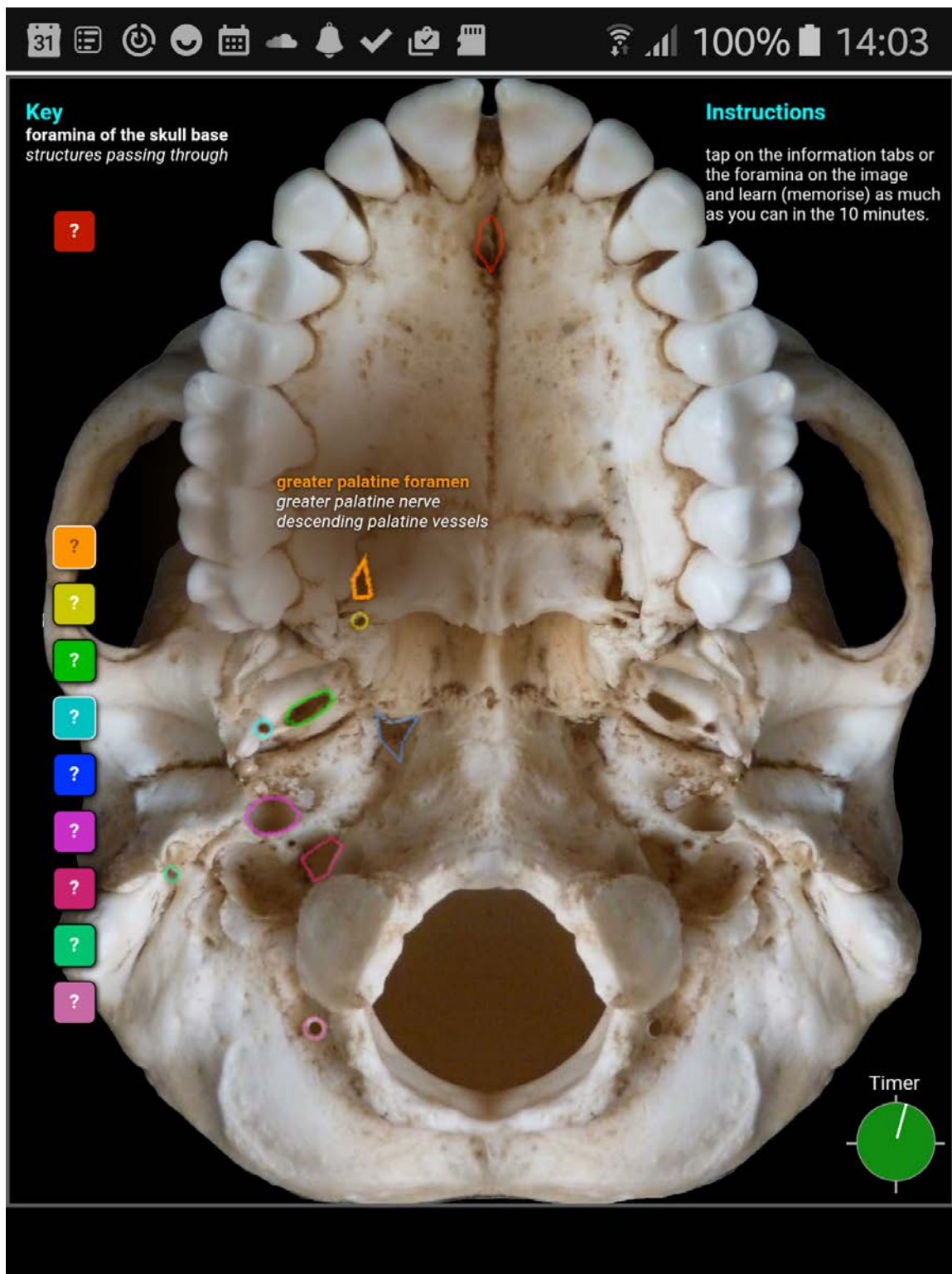
For the experimental group, screen 06 provided instructions on how to access the learning activity on a mobile device. The m-learning activity was created as a separate app that could be downloaded to a smartphone or tablet computer. A link was provided to access the application which was designed to only work on a mobile device. This prevented the participants accessing the mobile activity on a PC which would have invalidated the study. The m-learning activity was designed to be a web application which permits the participants to access it without having to install any additional software on their devices.

Mobile Application created for the m-learning activity

to be viewed on a smartphone or tablet:



Screenshot of m-learning activity web application opening screen on Samsung S7
smartphone (Samsung Inc.)



Screenshot of m-learning activity web application main screen

A detailed explanation of the application design is provided in a following section.

Screen 08 – NASA TLX explanation screen:

Screen 09 – NASA TLX self-report form:

Please rate each aspect of the learning task that you have just completed by moving the sliders below to indicate your choice from very low to very high. Scroll down to see all of the questions. Click the submit button. and wait for the green success message (scroll down to view) before clicking next.

Yes - Next

have you clicked the submit button and seen the green success message? You may need to SCROLL DOWN to the bottom of the form to see the success message

Rating the workload placed on you by the learning task

Six rating scales are displayed Below. Please evaluate the learning task that you have just completed by marking each scale at the point which matches your experience. Each line has two endpoint descriptors that describe the scale, move the slider to the right or left with the mouse until it points at the desired location. Please consider your responses carefully in distinguishing among the task conditions. Consider each scale individually. Your ratings will play an important role in the evaluation being conducted, your active participation is essential to the success of this experiment, and is greatly appreciated.

Mental Demand

How mentally demanding was the learning task?

How much mental and perceptual activity was required (e.g. thinking, deciding, calculating, remembering, looking, searching, etc) Was the task easy or demanding, simple or complex, exacting or forgiving?

Use the slider to indicate the amount of mental load you experienced.

VERY LOW VERY HIGH 50

Screenshot of screen 09a of the online data collection tool.

This screen features the NASA TLX questionnaire (Hart and Staveland, 1988) as an online version developed by the researcher. The scales are represented by dynamic sliders that can be manipulated by dragging with the mouse cursor. The position of the slider automatically generates a number (screen right) to represent the degree of load represented by the position of the slider. There was a descriptor provided for every dimension of the NASA TLX and the participants could scroll the document to reveal all of the sliders for each dimension. The screenshot above shows the first part of the NASA TLX and fig. 3.13 shows the second part of the test, designed to weight the final score by asking the participants to choose pairwise from each of the dimensions. This was facilitated by the use of an HTML feature known as radio buttons. These are designed to be used whenever the user is required to make an either/or choice as only one button can be activated at a time. Furthermore, the form used to collect the NASA TLX data was designed so that it could only be submitted if all of the required items had been completed, to avoid incomplete data submission by the participants.

Please rate each aspect of the learning task that you have just completed by moving the sliders below to indicate your choice from very low to very high. Scroll down to see all of the questions. Click the submit button, and wait for the green success message (scroll down to view) before clicking next.

Next

Sources of Workload

The evaluation you are about to perform is a technique developed by NASA to assess the relative importance of six factors in determining how much workload you experienced. The procedure is simple: You will be presented with a series of pairs of rating scale titles (for example, Effort vs. Mental Demand) and asked to choose which of the items was more important to your experience of workload in the task that you just performed. Each pair of scale titles are presented below. Select the Scale Title that represents the more important contributor to workload in the learning task that you just completed.

Example: If you thought that the amount of **time** allowed made the task more difficult than the amount of **physical effort** required you would select "**temporal**".

Of the two workload measures below, which one contributed most to the task you just completed?

☐ Mental Demand

☐ Physical Demand

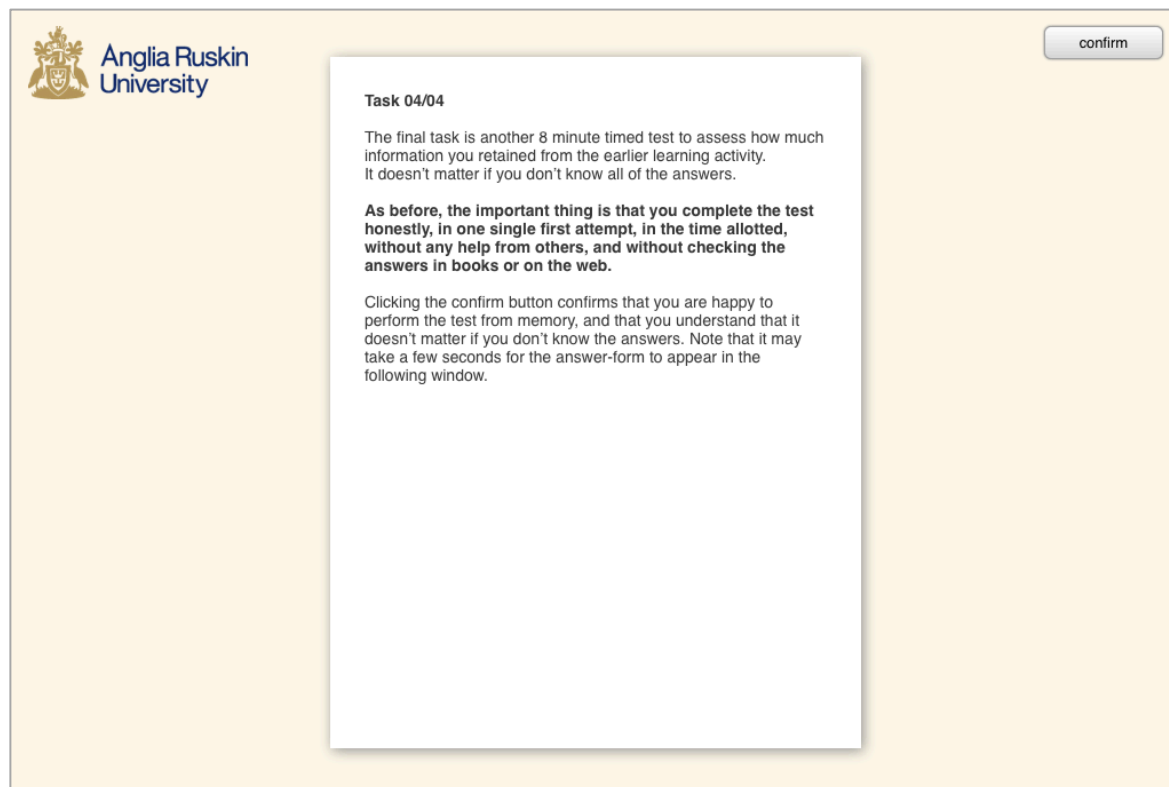
How much mental and perceptual activity was required (e.g. thinking, deciding, calculating, remembering, looking, searching, etc) Was the task easy or demanding, simple or complex, exacting or forgiving?

How much physical activity was required (e.g. pushing, pulling, turning, controlling, activating, etc)? Was the task easy or demanding, slow or brisk, slack or strenuous, restful or laborious?

Screenshot of screen 09b of the online data collection tool.

Having completed the NASA TLX and submitted the form, the online data collection tool automatically entered the submitted data into an online database and also sent an auto-generated email to provide a backup of the data. The participants were then required to click the next button to access the final part of the data collection process. The collection tool prompted them to ensure that they had submitted their NASA TLX form before continuing (see screenshot 9a, top left of the screen).

Screen 10 – Post-test explanation screen:



Screenshot of screen 10 of the online data collection tool.

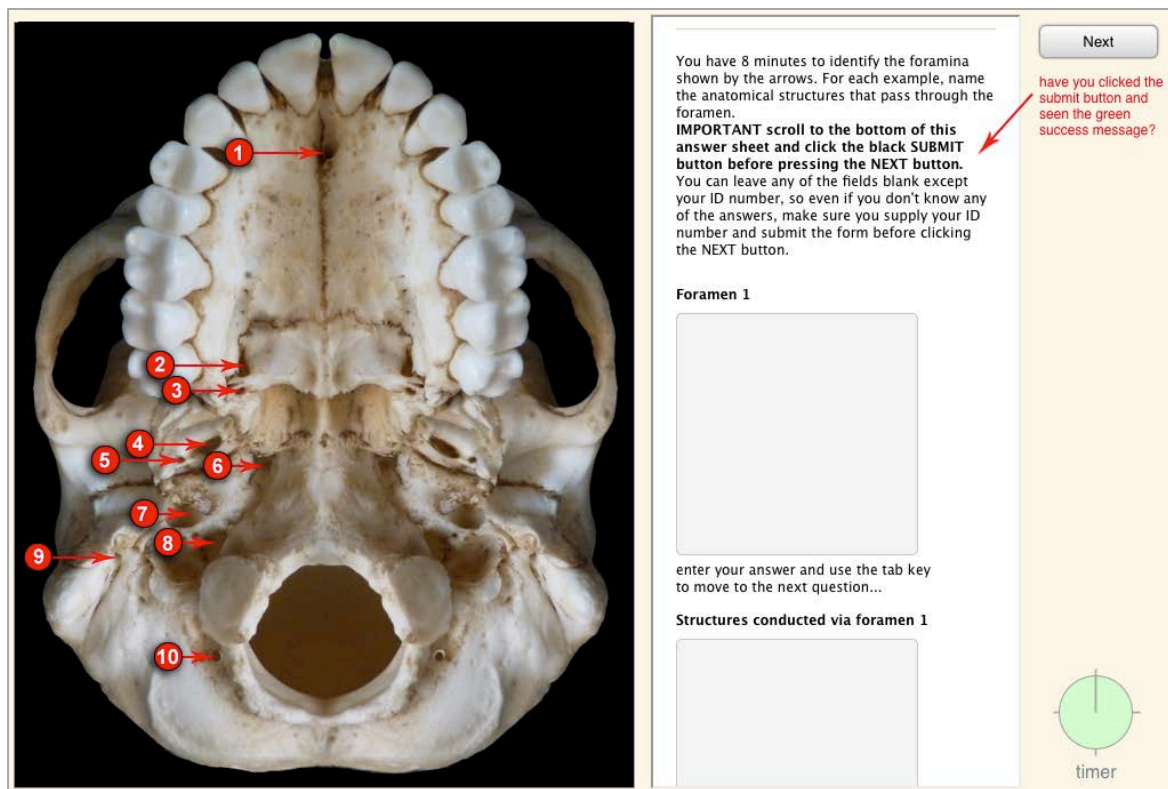
This screen explains that there will be a post-test and asks the participants to conduct the test from memory rather than by finding the answers in a book or online. There is a potential source for invalidation here because having read this screen the participants could theoretically spend more time learning the anatomical area from books or web sites. This was controlled for by the online database which time and date-stamped all of the content as it was submitted. In at least one case, there was a long period of many hours before a participant completed this form, suggesting that they had taken extra time to learn about the subject area before completing the post-test.

The rationale for giving the post-test after the NASA TLX rather than after the learning activity was two-fold:

Firstly, the items learnt would have had time to be committed to long term memory.

Secondly the data from the NASA TLX were considered to be more valuable to the study, so it was decided to acquire these first.

Screen 11 – Post-test on the Anatomy of the Skull Base:



Screenshot of screen 11 of the online data collection tool.

This screen gives a post-test to the participants to see how many of the anatomical features were learnt in the learning activity. The design is identical to the pre-test described in the earlier section.

Screen 12 – Task-completion screen



Screenshot of screen 12 of the online data collection tool.

The final screen reminds the participants that their data will be kept securely and that they will remain anonymous. The email address of the researcher is provided to allow any further queries to be made.

Appendix B : The NASA TLX Data Collection Form

NASA TLX Sheet 01

Mental Demand - how mentally demanding was the learning Task?

Very Low

NASA TLX Sheet 02

For each pair of workload measures below which one contributed most to the touch-screen learning task you just completed. Please circle just ONE from EACH pair.

Mental demand	Physical Demand
Performance	Mental demand
Temporal Demand	Effort
Performance	Temporal Demand
Performance	Frustration
Mental Demand	Effort
Temporal Demand	Mental Demand
Effort	Performance
Effort	Frustration
Effort	Physical Demand
Physical Demand	Temporal Demand
Temporal Demand	Frustration
Physical Demand	Performance
Frustration	Mental Demand
Frustration	Physical Demand

(Hart and Staveland, 1988)

Appendix C : Pre/Post-Test Data Collection Form

Activity 01 Pre-test

Gender (circle) M F

Age in years

Device Used (make and model)

	Identify (name) The Foramen	Identify (name) the structures conducted via each foramen. (name as many as you can)
1		
2		
3		
4		
5		
6		
7		
8		
9		
10		

Appendix D : Participant Information and Consent Form

My name is John Talbot I am a senior lecturer in Magnetic Resonance Imaging, research and applied anatomy at Anglia Ruskin University in Cambridge. I am currently undertaking a research project entitled:

“An Empirical Study to Assess the Impact of Mobile Touch-Screen Learning on User Information Load”.

Thank you for showing interest in taking part in my research project. Here are the answers to some questions you may have about participating.

Introduction

There is a new branch of learning, known as m-learning, that uses mobile phones and touch-screen computers as learning tools. It allows the student to study wherever they happen to be, and permits the use of images, sound and user interaction.

Previous research suggests that teachers sometimes design their teaching materials to suit the features offered by this new technology rather than to address the way students actually learn. This is of concern because meaningful learning may be hampered rather than enhanced by the use of the technology.

A research study into how these devices affect learning (for better or worse) would be therefore useful when designing learning activities.

What is the purpose of the research?

Previous studies have shown that certain types of learning activity require more mental effort from the learner. This is not necessarily because the subject is more difficult, it may be due to the way the materials are presented. My project is intended to compare two ways of presenting learning materials. I want to discover whether the degree of mental load is different for each method, and if so whether this affects the learning outcome. Did one method result in a lower mental load, and did this allow the student to learn more?

Do I have to take part in the study?

No, your participation is entirely voluntary and even though this data collection will take less than an hour of your time, you are still free to withdraw if at any point you do not wish to continue. I would like you to take part however, because I believe you can make an important contribution.

What will I be asked to do?

There are four stages to the task and they can all undertaken remotely (online):

- A pre-test to assess your prior knowledge of the learning topic.
- A timed (10 minute) learning activity.
- Another short survey to ascertain how much of a burden you felt the learning task placed on you. This will require you to rate six attributes of the activity using a scale from “very low” to “very high”.
- A post-activity test to ascertain what was learnt during the activity.

The entire process should take less than one hour of your time but it is essential that all five stages of the process are completed consecutively in one sitting.

Are there any reasons I should be excluded from the study?

People have been invited to take part in the study because they are interested in learning or distance-learning. This may be formal learning, or just an informal desire to learn. I intend to include a variety of ages and have representation from both male and female participants. This leaves the study open to most people, however, you must be able to fully understand written English in order to be able to read the questions and write your answers.

Are there any risks to me or discomfort in undertaking the task?

No, none are anticipated. Many people don't like taking tests, but the results of your test will be anonymous. You will not receive your results, and I cannot link you to your test result as you will not be required to provide any personally identifiable information.

I hope this will make you more comfortable about taking part. Remember you can choose to withdraw from the process at any time during the online data collection tasks. Simply close your internet browser.

What are the benefits of this research project?

The main benefit will be in informing the way distance learning materials are constructed for future students, or for people aiming to learn factual information online or via modern technology such as mobile phone apps. There will be no direct benefit to you, but there may be an indirect future benefit if you fall into the one of the groups mentioned above.

How is my privacy maintained?

With the approval of my supervisor I am not required to collect any personally identifiable information. All the information you provide will be kept safely on a secure password-protected non-networked data storage device and will only be accessed by myself as the researcher. No identifiable data will ever be released to a third party. The only exception to this would be if you provided data that resulted in you (or others) being put at risk of harm. This is most unlikely in the context of this study. In the event of the research findings being published or made public, only combined data will be included. You will not be identifiable. Although the data collection is undertaken using online computing, all data collection, storage and processing will comply with the principles of the Data Protection Act 1998 and the EU Directive 95/46 on Data Protection. No personally identifiable data will be stored online or in the "cloud". You can learn more about the Data Protection Act here:

Consent

Any information that you provide in the following surveys and tests may be used as part of a doctoral thesis and may subsequently be published.

All data collected will be stored securely, accessible only to me and my two academic supervisors, and are subject to the Data Protection Act safeguards. No information will be released or published that might identify you, or allow others to discover your personal answers or views that you may provide in the following exercises.

Participation is voluntary and you can withdraw from the process at any time during the data collection by discontinuing the data collection or keeping the data collection forms for your own disposal.

Completing the data collection activities and submitting the data collection forms signify your understanding and consent to the above.

Appendix F : HARDWARE AND SOFTWARE RESOURCES

Software

App Authorware:

Hype (Version 2.5.2. 338) (Tumult Software, 2015)

Rapidweaver (Realmac Software, 2015)

Adobe Photoshop (Adobe Systems Inc.)

Adobe Audition (Adobe Systems Inc.)

Hardware

App-authoring hardware:

iMac 3.4 GHz Intel Core i7 running OS X 10.9.3 (13D65) (Apple Inc.)

App testing hardware:

Apple iPad (Apple Inc.) running iOS 7,

Samsung Galaxy S2 Tablet (Google Inc.) running Android Lollipop.

Cube I5 tablet computer running Windows 8.1 (Cube Inc.),

Samsung Galaxy Ace, S5, Note 3, Note 4 and S7 Smartphones running Android KitKat,

Lollipop and Marshmallow OS.

iPhone 4, iPhone 5, iPhone 6, running iOS 7. (Apple Inc.)